The diversity puzzle

Mäs, Michael

THE DIVERSITY PUZZLE

EXPLAINING CLUSTERING AND POLARIZATION OF OPINIONS

The research of Michael Mäs has been supported by the Netherlands Organization for Scientific Research, NWO (VIDI Grant 452-04-351).

The graphic on the cover pictures the content of the first chapter if this book. The graphic on the back side is based on the Dutch summary. Both have been generated with *www.wordle.net*.

ISBN: 978-90-367-4491-1

Printing: Ridderprint Offsetdrukkerij B.V., Ridderkerk

RIJKSUNIVERSITEIT GRONINGEN

THE DIVERSITY PUZZLE
EXPLAINING CLUSTERING AND POLARIZATION OF OPINIONS

Proefschrift

ter verkrijging van het doctoraat in de
Gedrags- en Maatschappijwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
maandag 15 november 2010
om 14.45 uur

door

Michael Mäs
geboren op 10 mei 1981
te Weimar, Duitsland

Promotores:	Prof. dr. A. Flache Prof. dr. R.P.M. Wittek Prof. dr. W. Raub
Copromotor:	Dr. K. Takács
Beoordelingscommissie:	Prof. dr. A. Diekmann Prof. dr. M. W. Macy Prof. dr. F. N. Stokman

Acknowledgements

This book summarizes my research as a graduate student at the ICS in Groningen. I owe big thanks to the ICS. It creates a very stimulating environment for research and it gives young scientists an optimal framework to discover and develop their research interests. I am very grateful for the guidance and support from many ICS members.

First of all, I would like to thank Andreas Flache who was my main supervisor. He thought up an intriguing research project and guided me through it. However, he also supported my attempts to extend and modify the project. I am very much looking forward to our future work and tackling new puzzles together with Andreas.

I also received valuable and inspiring advice from Werner Raub, Rafael Wittek and Károly Takács who were part of the supervision team as well. I want to thank the members of the manuscript committee, Andreas Diekmann, Michael Macy and Frans Stokman, for their encouragement and comments.

During my PhD time, I shared office with several colleagues. Three of them have become close friends. I want to thank István Back, Katya Ivanova and Tobias Stark who made the last years very enjoyable. Furthermore, I want to thank everybody who took care of my two birds when I was not in Groningen.

I am grateful to Dirk Helbing who invited me to Zurich for an internship. This was a very inspiring time for me and has influenced my thinking about social processes a lot. In addition, I thank Dietmar Huber, Heiko Rauhut, Peter Felten, Sergi Lozano and Anders Johansson for making my visits in Zurich so rewarding. I also want to thank my teachers from Leipzig, Kurt Mühler, Karl-Dieter Opp and Thomas Voss.

My research benefited a lot from the research infrastructure of the Faculty of Behavioural and Social Sciences of the University of Groningen. In particular, I want to thank the concierge service and the ICT support team who made it possible to conduct extensive computer simulation studies and helped me with organizing the laboratory experiments.

I am also very grateful to my parents, grandparents, my brother and his family and my girlfriend's family. Thank you for your support and visiting me in Groningen.

Conducting fundamental research and writing a PhD thesis can be a very frustrating and painful process. The most difficult aspect of the last 5 years, however, was the fact that I lived separated from my girlfriend, Mandy Weidner. Her love, trust and support made this research possible.

Contents

I. Overview of the book

I.1. Social-influence processes in human groups

Example 1: Reaching consensus in the jungle.

Your plane crashed in the jungle. When the plane hit the ground, all crew members were killed but you and three other passengers survived. The group of survivors is demographically mixed. There are two males and two females. Two of you are black and two are white. After having waited for days on end in the jungle, you still have not been rescued. The four of you discuss what to do and agree that the group should try to find a way out of the jungle. The problem is that no one has a clue in which direction to go. One of the other passengers asks you: "Listen, as a sociologist you should be an expert on this. Is it wise to discuss the problem or should one of us decide for the group? To maximize our chance of survival, we have to remain together as one group. Finding a consensus would be optimal because all of us would support the decision. However, when we discuss the problem as a group, people might disagree and someone might pick a fight. This is the worst thing that can happen in our situation. Alternatively, one of us may assume the leadership and decide for the group. This would be better than a conflict but certainly worse than the consensus." What would your advice be? How likely is it that in a discussion all the survivors can be convinced that one given direction is the most promising? How would the group's demographic diversity affect the discussion? Will it make conflicts more or less likely?

Example 2: The Oscars.

Every year, the Academy of Motion Picture Arts and Sciences nominates five actors for the Oscar for the Best Actor for that year. On the night when the awards are presented, everybody who works in the movie business is in Hollywood. When the winner is announced, all these people will have an opinion about this decision and it is likely that opinions will vary considerably.

Some will believe that the winner very much deserved the award. Others will feel that the winner's work is overrated and another nominee deserved the award more. Later, at the famous Oscar parties, people will talk about the Academy's decision. They will discuss why they think that the winner deserves the Oscar or why not. As a result of these discussions, will the opinions of the movie people converge? In other words, will there be consensus on whether the winner deserves the prize or will opinions remain different?

Example 3: How cool is rap music?

At the beginning of high school, a student becomes involved in rap music. Her class mates are interested in music but do not have a clear-cut musical preference yet. How will the musical preferences of the other students develop in the subsequent months? Will the rap fan infect the others? Under what conditions will the others develop different musical preferences?

These examples have two central aspects in common. First, each member of the various groups holds an *opinion* on a certain issue. In other words, each of them has evaluated a certain object (a certain direction to go, the decision of the Academy, rap music). Opinions vary between two extremes and can be measured on a metric scale. For example, the opinion of each plane-crash survivor can be quantified using the degrees on the compass rose. Each survivor prefers a specific direction. The more degrees that the preferred direction of another survivor differs from this direction, the more the two actors disagree. Second, the members of each group interact and may *influence* each others' opinions. In particular, they may persuade each other and adjust their opinions in such a way that they become more similar.

Research shows that social influence plays an important role in social interaction. It is argued, for instance, that interaction partners exchange arguments and persuade each other that certain opinions are more adequate (Myers 1982; Wood 2000). Other theorists argue that interaction partners may exert social pressure to conform with each other (Festinger, Schachter and Back 1950; Homans 1951). Furthermore, cognition theories (Festinger 1957; Heider 1967) imply that we want to be similar to people we like to interact with. To achieve this, we might try to convince our friends to adopt opinions and behavior similar to ours, or, the other way around, we might change our opinions and attributes to conform to those of our friends. Furthermore, social influence may also result from imitation (Akers et al.

1979). It has been argued, for instance, that in situations of high uncertainty it can be rational for individuals to imitate the behavior and opinions of others (Bikhchandani, Hirshleifer and Welch 1992).

Empirical studies also demonstrate the importance of social influence. For instance, psychological experiments consistently show that subjects adjust their opinions after having been informed about the opinions of another person (see Wood 2000 for an overview). In particular, when subjects know that they share some attribute with that person, they tend to decrease their opinion distance (e.g. Berscheid 1966; Sampson and Insko 1964; van Knippenberg and Wilke 1988). In addition, research has been conducted on the outcomes of discussions in social groups. In these experiments, typically groups of four subjects meet and discuss some issue (Johnson and Johnson 1982). It has been found that, as a result of the discussion, these participants adjust their opinions (Myers 1982; Wood 2000).

This book is concerned with explaining the opinion dynamics that social influence causes in social groups. Based on these explanations, we seek to predict the distribution of opinions which result from social influence processes. Under what conditions will the members of a group find a consensus? Under what conditions will initial opinion differences persist? Is it possible that opinion differences between subgroups will increase over time? Is opinion consensus always stable, or can homogeneous groups split up into subgroups with opposing opinions? Our goal is to develop theories which can be applied to all social groups, from small groups such as work teams and school classes to large scale groups such as social networks on the Internet.

Our focus in this book is on finding an *explanation* for social influence dynamics. We developed theories to describe opinion dynamics that result from social influence and to generate predictions about the outcomes of influence processes. It has been argued that theories of social-influence may also help us understand under what conditions groups find the best or the correct answer to a given problem (see for instance Kennedy 1998). It has also been proposed that management teams arrive at better decisions than individual managers (Johnson and Johnson 1982). Following a similar line of reasoning, it has been claimed that there is something called the *wisdom of the crowds* (Surowiecki 2005) in the sense that aggregating the information of crowd members results in better decisions than the decisions from single crowd members. That particular topic, however, is beyond the scope of this book. Before we study under what conditions groups make good decisions, we need to have a reliable theory of opinion formation in groups. We will show in the following section that existing models have several shortcomings and that therefore our focus will be on elaborating on the existing theories.

I.2. Existing theories of social-influence dynamics and their shortcomings

In this section we will review the existing theories which have been developed to understand social-influence dynamics. We will summarize the answers that these theories give to the questions we have just formulated and point to their shortcomings. Finally, we will formulate two research questions which the existing models fail to answer.

I.2.1. Classical social-influence models

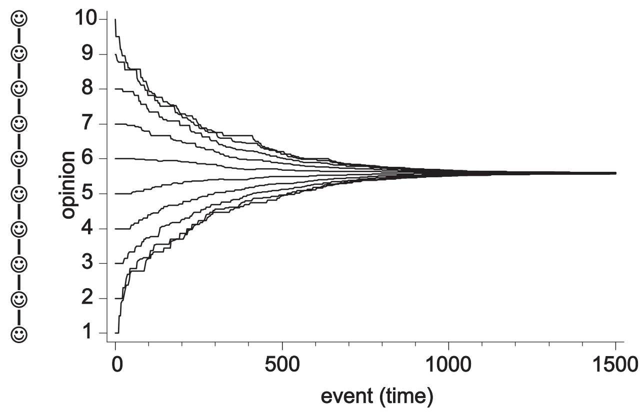
Already in the 1950s, researchers have developed formal theories of social-influence processes in groups (Abelson 1964; Berger 1981; DeGroot 1974; French 1956; Friedkin and Johnsen 1990; Harary 1959; Lehrer 1975). These models express the opinion of each group member as a real number which can vary between two extremes. As an example, this value might reflect the opinion on rap music of one of the students from Example 3. Small values would indicate a negative opinion and high values a positive opinion.

In contrast to the classical literature on social-influence dynamics, several models have been developed which assume that actors influence each other on *nominal* variables (Axelrod 1997; Carley 1991; Galam 2005; Liggett 1985; Mark 1998; Mark 2003; Nowak, Szamrej and Latané 1990; Sznajd-Weron and Sznajd 2000). These nominal variables represent for instance whether the actors adopt a piece of information (Carley 1991) or the political party the actors will vote for (Liggett 1985). However, many of the attributes which are open to social influence are better described using metric scales. For instance, very few people agree or disagree exactly with the program of a political party. Instead, people usually agree to a certain degree. We will show in this section that existing social-influence models are problematic when they are applied to such continuous dimensions. Our aim is to overcome these problems. We will therefore stick to the assumption of the classical social-influence models and focus on continuous opinions. We should also point out that nominal variables can not serve as an approximation of metric attributes. This is because predictions of social-influence models critically depend on whether actors influence each other in terms of nominal or continuous dimensions (Flache and Macy 2006a).

When incorporating the social-influence mechanism, modelers assume that the actors adjust their opinions in such a way as to become more similar to the other group members.

In particular, they assume that actors adopt an opinion equal to the weighted average of the others' opinions. As a result, models take into account the fact that not all group members are equally influential for an actor. The impact that an actor *j* has on actor *i*'s opinion is expressed by a weight $w_{i,j}$. A weight of zero represents the fact that *j* has no influence on *i*'s opinion. Values above zero indicate how strong the impact of *j* is on *i*. Classical models assume that weights do not change over time.

Figure I.1: Opinion trajectories of 10 actors in the classical social-influence model



To illustrate the opinion dynamics which follow from these assumptions, we show in Figure I.1 a typical social-influence process in a group of 10 actors, named 1 thru 10. At the outset of the influence dynamics, each actor holds an opinion the same as his name. Furthermore, the group has a very simple influence network: a line. That is, Actor 1 is only influenced by Actor 2 ($w_{1,2}$=1). Actor 2 is only influenced by Actors 1 and 3 ($w_{2,1}$= $w_{2,3}$=1). Actor 3 is only influenced by Actors 2 and 4 ($w_{3,2}$= $w_{3,4}$=1) and so on. The remaining weights (e.g. $w_{1,3}$) are zero. For the sake of illustration, we have modeled the group's social-influence process as a sequence of events. At each event, we picked one of the ten actors and adjusted his opinion. In line with classical social-influence models, we assigned a new opinion that was similar to the average opinion of the actor and his network partners.

Figure I.1 demonstrates that the opinions of the actors converge and become basically identical in the long run. At the beginning of the process, Actors 2 thru 9 do not change their opinions because the influences from their two contacts balance each other out. However, Actors 1 and 10 are influenced by only one group member. In both cases, the opinion of the contact is less extreme than the actor's own opinion. As a consequence, Actors 1 and 10 develop more moderate opinions. This, in turn, affects the opinions of Actors 2 and 9. Because their extreme contact (1 and 10, respectively) has developed a more moderate opinion, the opinions of 2 and 9 also shift towards a less extreme view. These changes, in turn, trigger opinion adjustments for Actors 3 and 8 and so on.

It has been shown that the result of the example from Figure I.1 can be generalized. More precisely, researchers have proved analytically that classical social-influence models *always* predict convergence of opinions towards global uniformity unless there is a subset of actors that is completely cut-off from outside influences (Abelson 1964; Berger 1981; DeGroot 1974; Harary 1959; Lehrer 1975; Wagner 1982). Thus, irrespective of the structure of the influence network and the initial opinion distribution, consensus appears to be unavoidable. According to the classical models, consensus formation can only be stopped by social separation.

This result is puzzling because it contrasts with the diversity of opinion that we observe in real life. For instance, research on group dynamics in work teams demonstrates that team members often fail to decrease their differences of opinion within the team (van Knippenberg and Schippers 2007). Members of work teams interact frequently and are therefore easily able to influence each other. Yet, differences of opinion do not always decrease; sometimes they even intensify (Early and Mosakowski 2000).

In a similar way, long-term studies on the social attitudes of Americans show that there is significant opinion variation on relevant issues such as feelings towards blacks, views on gender equality and the evaluation of abortion (DiMaggio, Evans and Bryson 1996). Studies did find decreasing variance for many opinion dimensions (e.g. opinions on matters related to crime and justice). However, studies also found that the variance of several opinion dimensions has *not* changed significantly from the beginning of the 1970s to the beginning of the 1990s (DiMaggio, Evans and Bryson 1996; Mouw and Sobel 2001). A study which included data up until the year 2000 found that variance has actually increased for dimensions such as sexual morality and attitudes towards poor people (Evans 2003). There is also research showing that the American diversity in terms of lifestyles and consumption tastes has increased (for an overview: Fischer and Mattson 2009).

Social separation does not provide a plausible explanation for these findings. Modern societies have developed very efficient transportation and communication technologies which make it very easy to be in contact even with geographically distinct people (Greig 2002). Furthermore, modern "worlds" are surprisingly small (Milgram 1967). It has been demonstrated that in modern societies most pairs of individuals seem to be connected by a very short path through the network (Barabasi 2003; Travers and Milgram 1969; Watts

2003; Watts and Strogatz 1998). This suggests that in modern societies individuals are exposed to influence from various directions. Yet, diversity of opinion remains high.

An alternative explanation for opinion diversity may be that some opinions are related to material interests. For instance, rich people typically hold negative opinions on issues like inheritance tax increases. Such interests may interfere with social-influence processes and bring opinion convergence to a stop when actors refuse to adjust their opinions because this would contradict their material interests. In line with this argument, empirical research in the US has found opinion differences between respondents with a college degree and respondents with no more than a high-school education (DiMaggio, Evans and Bryson 1996). However, this research also showed that opinion differences between the two groups decreased between the 1970s and the early 1990s. Furthermore, these differences can not explain opinion diversity on issues that are unrelated to material interests such as sexual morality (Evans 2003) and the attitude towards abortion (DiMaggio, Evans and Bryson 1996).

A related point has been made by Friedkin and Johnson (1997) who assumed that actors seek to adopt opinions which mirror their views at the outset of the influence process. Their social-influence model is able to generate opinion dynamics with decreasing opinion variance at the beginning of the process but stable opinion diversity in the long run. Dynamics reach equilibrium when social pressures to conform to others' opinions and the striving towards the initial opinion cancel each other out such that opinions remain stable. We do not question this assumption. However, we would like to point out that this model can not generate the *increasing* opinion diversity which empirical studies have found on issues such as sexual morality and attitudes towards poor people (Evans 2003). Moreover, the Friedkin-Johnson model can generate social-influence dynamics that result in opinion diversity only in those settings that are already at the outset of the influence process characterized by high opinion diversity. This model does not, however, offer an explanation for the emergence of opinion diversity. In other words, the model predictions rely on the existence of opinion distributions which the model cannot generate endogenously. In this book, we will develop models which do not have these shortcomings.

In summary, classical social-influence models predict that opinions will converge over time. However, there is a conflict between this prediction and the opinion dynamics which we observe empirically. Puzzled by this, Abelson wondered already in 1964 "what on earth one must assume in order to generate the bimodal outcome of community cleavage studies?"(153). With this, Abelson identified a puzzling and tenacious research problem. In this book, we will follow Abelson's lead and seek to explain why there is persistent opinion diversity despite social influence.

I.2.2. One potential solution: Selection of similar interaction partners

One potential solution to Abelson's puzzle has been proposed in the literature on cultural and group dynamics (Axelrod 1997; Carley 1991), in the literature on decision-making in evolving networks (Stokman and Zeggelink 1996a; Stokman and Zeggelink 1996b) and in the social-influence literature (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002). The assumption found here is that actors tend to interact with those others who hold similar opinions and avoid influence from dissimilar others.

This notion is supported by two prominent theories (Byrne 1971). First, the reinforcement approach (Byrne 1961) argues that humans do *not* have a general predisposition for interacting with similar others. However, whenever we interact with similar others we feel rewarded. When an interaction partner "offers us validation by indicating that his percepts and concepts are congruent with ours, it constitutes a rewarding interaction and, hence, one element in forming a positive relationship" (Byrne 1961: 713). At the same time, interaction with dissimilar others constitutes a negative stimulus because we learn that our opinions may be wrong. Furthermore, such interactions may end up in punishing discussions and conflicts. In other words, during interaction with similar and dissimilar others we learn that meeting with similar others just feels better and, hence, we start to prefer such interaction partners.

A second explanation is based on cognitive theories. These theories argue that humans strive for balanced cognitions (Festinger 1957; Heider 1967). This is achieved when we have positive emotions towards persons who are similar to us, and when we have negative emotions towards persons who are dissimilar. For instance, the two cognitions "I *dis*like Tom" and "I agree with Tom" are unbalanced. According to these cognitive theories, unbalanced cognitive constellations are unpleasant for us and we try to avoid them. In the example, this can be achieved by changing our relationship to Tom from disliking to liking. In short, when we learn that an interaction partner holds similar opinions, it just feels good to like him.

Empirical research also supports the assumption that individuals tend to interact with similar others. Social psychological experiments along the Similarity-Attraction-Paradigm (for a comprehensive summary see Byrne 1971) support the fact that humans evaluate similar others more positively than dissimilar others. In these experiments, subjects are introduced to a stranger along with information on the stranger's opinions. Experiments show consistently that subjects rate strangers more positively the more they agree with the opinion of the stranger. In addition, research on the determinants of friendships shows that humans tend to nominate similar group members as their friends. To illustrate, it has been found that students prefer friends with similar smoking behavior (Mercken et al. forthcoming). Similar effects have also been found for delinquency (Burk, Steglich and Snijders 2007) and alcohol consumption (Pearson, Steglich and Snijders 2006).

Researchers have included the notion that individuals tend to select similar others as interaction partners in social-influence models. In particular, it is assumed that actors refuse to be influenced by actors who are too dissimilar. The maximum dissimilarity that actors consider acceptable is called the confidence level. Accordingly, such models are called *models with bounded confidence* (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002; Lorenz 2007). To incorporate this notion into the classical social-influence framework, it is implemented that the influence weights are a function of the opinion similarity between the actors. In particular, weights are assumed to be zero as the dissimilarity between two actors exceeds a certain threshold; otherwise, they are one[1] (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002).

The interplay of the selection of similar others and social influence creates a feedback process which can lead to the development of distinct clusters of similar actors (Axelrod 1997; Carley 1991; Hegselmann and Krause 2002). In this process, similar actors interact

---

[1] To be more precise, models of decision making in evolving networks (Stokman and Zeggelink 1996a; Stokman and Zeggelink 1996b) are based on a slightly different implementation of the bounded-confidence assumption. In these models, actors influence each other only if they have established a mutual network relationship. Furthermore, the probability that an actor will attempt to create such a relationship and that this attempt will be reciprocated depends on the opinion similarity between the two actors. It is thus *un*likely that dissimilar actors will establish a mutual network relationship. Furthermore, these models assume that, if an actor *i* attempts to create a relationship but the potential network partner *j* rejects this offer, then the probability that *i* will repeat the attempt to create a network relationship to *j* at a later time decreases. Since rejecting network ties is more likely the more the opinions of *i* and *j* differ, the probability that dissimilar agents will create a relationship very likely decreases quickly and eventually adopts zero. This rules out any future influence between these actors.
Note that some models of decision-making in evolving networks (Stokman and Zeggelink 1996b) combine the bounded-confidence assumption with an additional mechanism that may prevent opinion convergence even in a connected network. Contrary to the classical social-influence models and the models which we will develop in the following chapters, some models of decision making in evolving networks assume that opinions can adopt only one of two values, zero or one. In addition, it is assumed that actors adopt an opinion which is similar to the median of all network partners' opinions (Stokman and Zeggelink 1996b: page 401 and footnote 15). As a consequence, an actor *i* may have a relationship to an actor *j* who holds an opinion value of one. However, if there is a sufficient number of other network partners who hold the opposite opinion (zero), then *i* will not be influenced by *j* and will adopt an opinion of zero.

and thereby become even more similar. This, in turn, leads to further interaction which again increases similarity. In this process, distinct clusters of similar actors can emerge. Actors at the center of a cluster pull actors from the border of their cluster closer to themselves. In doing so, they pull them away from other clusters. Subsequently, the members of a cluster become more and more similar and collectively ignore actors outside their cluster.

To illustrate how the interplay of selection and influence triggers cluster formation, Figure I.2 shows a scenario which we have replicated from Hegselmann and Krause's computer simulation study (2002). In this simulation, we modeled the opinion dynamics of 100 actors that hold random opinions at the outset. We imposed that the influence weights adopt the value one for pairs of actors with a dissimilarity smaller than 20% of the width of the opinion scale and zero otherwise.

Figure I.2: Opinion trajectories of 100 actors generated by the interplay of selection and social-influence (BC-model of Hegselmann and Krause (2002))

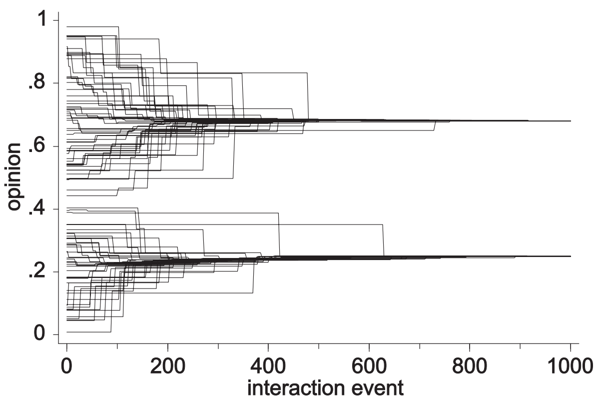


Figure I.2 shows that initially opinion is uniformly distributed. In contrast to the classical social-influence models (see Figure I.1), however, two clusters form. The dissimilarity between the clusters exceeds the imposed dissimilarity threshold of 20%. As a consequence, actors that belong to distinct clusters ignore each other. Dynamics settle when the clusters are internally perfectly homogeneous.

As Figure I.2 illustrates, the feedback process generated by the interplay of selection and influence can lead to opinion clustering and can stop the inevitable march towards consensus that classical social-influence models predict. However, we argue that this approach relies heavily on two problematic assumptions. First, the bounded confidence assumption proposes that actors refuse to interact with others who are too dissimilar. We do not question this assumption in general but it may be too strict to assume that dissimilar actors never ever interact. In Chapter V (see Figure V.1, Panel B) we show in more detail that including even an extremely small probability of interaction between actors who are too dissimilar to interact otherwise suffices to reactivate the march towards consensus of the classical models. As an illustration, let us take a look at the opinion distribution at the end of the process shown in Figure I.2. There are two distinct clusters which are sufficiently dissimilar to rule out further influence. But, if an actor happens to be influenced by the opposing cluster, this actor will move slightly nearer to this cluster. Subsequently, the actor will influence the other members of his own cluster and thereby pull them nearer to the opposing group. If such events occur more often, the clusters will gradually move nearer to each other and eventually be similar enough to exert influence on each other and merge. In real life, such events may seldom occur. However, if they can not be ruled out completely, then bounded-confidence models predict opinion convergence, too.

A second shortcoming of the bounded-confidence model is that it can generate clustering only when there is sufficient opinion variation already at the beginning of the influence process. In Figure I.2, for instance, the dynamics started with a uniform opinion distribution. Such populations are comprised of extreme actors who will pull others towards them. This is crucial because without a sufficient number of extremists, emerging clusters will be too similar and therefore merge in the long run. This problem is rooted in the nature of social influence. It implies that interaction partners become more similar. It follows that diversity can only decrease over time. Thus, there is no endogenous mechanism that can create diversity. The bounded-confidence explanation of opinion diversity is problematic because it hinges on the existence of sufficient initial diversity, while at the same time the model fails to generate that diversity.

In sum, researchers have included the assumption that actors will tend to select similar interaction partners and refuse to interact with group members who are too different in their opinions. We argue that this is in fact a promising strategy for two reasons. First, this assumption is supported by empirical research. Second, it has been shown that the resulting models are capable of explaining opinion diversity. However, the new models fail at the same time to explain opinion diversity when only very small perturbations from the bounded-confidence assumption are included. This, and the fact that the models fail to create diversity, demonstrates that the bounded-confidence models still do not yet offer a satisfactory explanation for opinion diversity.

I.2.3. Research Questions

This brief review of existing social-influence models has identified a conflict between the models' predictions and empirical evidence. One the one hand, existing social-influence models imply that social influence triggers convergence cascades which lead to decreasing opinion diversity and eventually end up with perfect uniformity. Empirical evidence, on the other hand, does not confirm these convergence tendencies. On the contrary, opinion diversity often remains stable and can actually increase over time.

In this book, we have sought to overcome this conflict. In particular, our goal was to develop theories of social-influence dynamics which are capable explain two specific phenomena. First, following Abelson's lead, we looked for an explanation for *bimodal* opinion distributions. In other words, we sought to explain the parallel persistence of several (at least two) distinct but internally homogeneous subpopulations of actors. Explaining opinion clustering is important because there is empirical research showing that opinions tend to cluster for example within geographical regions (Glaeser and Ward 2006), sociodemographic groups (Mark 2003) and online communities (Lazer et al. 2009). We therefore reformulated Abelson's clustering puzzle as the first research question of this book:

> **Research Question 1:**
>
> How can the persistence of clustering despite social influence be explained?

Existing social influence models fail to generate another phenomenon. As Figure I.2 shows, bounded-confidence models may be able to generate clusters, but the distance between the clusters remains constant over time. There is, however, empirical evidence showing that opinions may polarize in the sense that subgroups with increasingly distinct opinions form. Polarization tendencies have been found in studies on the American public opinion on issues such as abortion, sexual morality and the war in Iraq (DiMaggio, Evans and Bryson 1996; Evans 2003; Fiorina and Abrams 2008; Fischer and Mattson 2009; Levendusky 2009). Furthermore, there is research on conflicts in work teams showing that task and emotional conflicts can intensify over time (Early and Mosakowski 2000). The fact that standard social influence models fail to explain polarization leads to the second research question:

Research Question 2:

How can the polarization of opinions despite social influence be explained?

I.3. Our answers to the research questions

Compared to the three examples that we introduced at the beginning, the existing social influence models appear to be very abstract. For instance, the models ignore possible demographic differences between individuals (see Example 1). What is also neglected is that social-influence may work differently for high-school students (Example 3) than it does for adults (Example 2). High abstraction is a desirable feature of a theory for two reasons. First, abstract theories have more analytical power in the sense that they can be applied to a wider range of empirical realms. For instance, a theory which is perfectly tailored to the opinion dynamics of survivors of a plane crash but can not be applied to other scenarios is not very informative. Furthermore, the more specific a theory is the fewer real-life situations it can be applied to. This makes it difficult to empirically test theories that are tailored to very specific phenomena (Popper 1972; Popper 1959). The second advantage of abstraction is that abstract theories make fewer assumptions. This makes theories simple and relatively easy to understand.

The existing social influence models are analytically powerful and they are simple, but they are also empirically inaccurate. We have shown, for instance, that these models fail to predict opinion polarization and opinion clustering. To overcome these shortcomings, we will need to forgo some of the analytical power and simplicity of the existing social-influence models and add new ingredients to the models. In doing so, we will follow the principles of Lindenberg's method *of decreasing abstraction* (1992). According to this method, one gradually decreases abstraction by adding new assumptions in a step-by-step process. Thus, instead of simultaneously adding several assumptions which potentially help to overcome empirical inaccuracy, the researcher adds one assumption after the other. After having included a new ingredient, one first tests whether the model changes do indeed add empirical accuracy. In other words, one tests whether including the new assumption helps in overcoming the shortcomings of the original model. This is central because adding assumptions which fail to increase the quality of the theory's predictions is a waste of analytical power and theoretical simplicity.

Furthermore, it is sensible to include new ingredients in the original theories independently one by one. This is because two ingredients may affect the predictions of a theory in a similar way. In this book, for instance, we discuss two different mechanisms, both of which can generate opinion polarization. However, we do not include both of these mechanisms in one model. To keep the number of assumptions in our theories low, we developed two competing models and studied under what conditions the two models predict polarization. We then identified conditions where the two models imply different predictions. These different predictions can then be tested empirically so as to test under what conditions each model implies more accurate predictions.

Finally, we want to point out that we mainly included *individual-level mechanisms* in the existing models. Individual-level mechanisms make assumptions about the behavior of the individuals who constitute the studied population. The social-influence assumption is an example of such a mechanism. In contrast, *structural factors* point to characteristics of the population. For example, it has been argued that opinion dynamics are affected by the existence of extreme opinion leaders (Hegselmann and Krause 2002), competing political parties (Fiorina and Abrams 2008) or media (Watts and Dodds 2007).

We do not question the fact that structural factors can have a significant impact on opinion dynamics. However, we do argue that individual-level mechanisms should be included in a theoretical model before structural factors are considered. This allows for testing whether a given research question can be answered even without including structural factors. For instance, can one explain polarization even in the absence of opinion leaders and political parties, or are certain structural factors necessary conditions for opinion polarization?

A second reason why we focused on individual-level mechanisms is that the effects of structural factors can crucially depend on which individual-level mechanisms are included in a model. In this book, for example, we study the effect of demographic diversity (a structural factor) on opinion polarization with two different models of individual behavior. In Chapters II and IV, we show that the two models have different implications concerning the conditions for this effect. To be able to understand why the two models imply different predictions, it is important to understand the dynamics that each individual-level mechanism generates. This is easier when structural factors are excluded.

In this book, we discuss four ingredients which we have added to the existing social-influence framework. We included three different individual-level mechanisms and one structural factor. In each chapter, we focused on one or two of these ingredients and studied whether they help answer the research questions. In the following, we will sketch how each ingredient might help in answering the two research questions. However, each of the ingredients has shortcomings, too. We will address these shortcomings in Section I.6.

I.3.1. Negative influence

Classical social-influence models are based on two central assumptions. First, it is assumed that individuals tend to interact with others who hold similar opinions (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002). Second, during interaction individuals influence each others' opinions in such a way as to become more similar (Abelson 1964; French 1956). In their review of the social-influence literature, Mason, Conrey and Smith (2007) recently recommended adding the negative counterparts of these two social mechanisms. In short, this suggests adding two further assumptions. First, individuals with very distinct opinions may not only refuse to interact but may actually dislike each other (Byrne, Clore and Smeaton 1986; Chen and Kenrick 2002; Pilkington and Lydon 1997; Rosenbaum 1986a; Rosenbaum 1986b; Smeaton, Byrne and Murnen 1989). Second, individuals may tend to increase the opinion distance with the disliked others. This notion has been called *negative influence* and has recently been included in social-influence models (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Mark 2003; Salzarulo 2006).

Negative influence could lead to polarization and might thus offer an answer to our second research question. How does it do so? Consider a population where the opinions are normally distributed at the outset. This population is comprised mainly of moderate actors but also a few extremists. The extremists from the opposing poles are highly dissimilar and will influence each other negatively. If this negative influence is strong enough, the extremists may thus become even more extreme. This can have consequences for the moderate actors because they are influenced positively by those extremists who are closer to them. When some moderates become more extreme, this can start a bandwagon effect in which the number of moderates decreases gradually until the population is split up into two maximally dissimilar groups.

I.3.2. Persuasive Arguments

The classical social-influence models assume that actors adopt opinions similar to the *average* opinions of their interaction partners. However, research on the outcomes of group discussions suggests that this assumption may be too simple (Myers 1982; Vinokur and Burnstein 1978). During interaction, individuals not only inform each other about their opinion, but they also talk about the arguments which they base their opinions on. Furthermore, it is possible that two actors hold identical opinions but base them on different arguments. When they interact and exchange arguments, they will provide each other with new reasons for their opinions. This could intensify their opinions in the sense that they develop more extreme views.

In combination with the selection of similar interaction partners, this elaboration of the influence process could explain opinion polarization. When individuals mainly interact with others who hold similar opinions they mostly exchange arguments which will intensify each others' opinions. As a consequence, also actors with rather moderate opinions would tend to become more extreme over time. This could aggregate to polarization.

For example, such a process might lead to opinion polarization during the Oscar parties (see Example 2). After the award ceremony ends, people will decide to which party they will go to. This decision will likely correlate to their opinion about the Academy decision. The friends of the winner and his colleagues will be very happy with the decision and will want to celebrate with the winner. Those, however, who do not have a close relationship with the winner but have a close relationship with one of the other nominees will likely hold more negative opinions. These people may not want to attend the party of the winner but may prefer to meet the losers to cheer them up. The consequence of these decisions is that the participants in a specific party will likely hold similar opinions. When people then talk about the decision of the Academy, they will likely provide each another with new arguments which support their initial opinions. As a consequence, initial opinion differences might intensify during the Oscar night.

I.3.3. Striving for uniqueness

Example 3 is concerned with the opinions towards rap music in a class of high-school students. These students are adolescents and, thus, are in a phase where they are very open to influence from peers. Hence, one would expect that after a few weeks all classmates will have adopted the opinion of the rap fan. On the other hand, adolescents also seek to define a *unique* identity for themselves. As a consequence, they might want to disagree with the majority opinion in their class. This suggests that students might want to deviate from an emergent opinion consensus and that consensus will be very unstable. We argue that the interplay of peer influence and striving for uniqueness can lead to development of several homogeneous subgroups with different opinions. Such constellations will be relatively stable because students will feel unique as they perceive opinion differences with many classmates. At the same time, there are also classmates with similar opinions. This will satisfy the students' need to be similar to peers.

Prominent sociological theories of social differentiation (Durkheim 1973; Durkheim 1982 [1895]; Simmel 2004 (1858); Turner 1995) suggest that striving for uniqueness is a strong force in modern societies. These theories argue that population growth and the development of new technologies initiate competition for resources. This, in turn, forces individuals to distinguish themselves from the mass. As an example, individuals specialize in their occupations in order to differentiate themselves from competitors. But in addition to this functional differentiation, individuals also want to differentiate themselves in terms of cultural dimensions and develop individual opinions and values. This idea is supported by recent research on the need for uniqueness (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980) which found that individuals tend to deviate from behavioral regularities.

Moreover, Brewer's optimal distinctiveness theory (Brewer 1991) and Social Identity Theory (Tajfel and Turner 1986) hold that there is a "fundamental tension between human needs for validation and similarity (on the one hand) and a countervailing need for uniqueness and individuation (on the other)"(Brewer 1991: 477). However, even though these theories capture opinion dynamics, research along these approaches has mainly focused on how humans self-categorize. In other words, attributes such as opinions are often assumed to remain unchanged when individuals feel too similar or too dissimilar to others. Instead, the research focuses on which set of attributes humans consider important when they define their identity. We went beyond these approaches and explicitly included the possibility that individuals may adjust their opinions in order to decrease tension between the needs for similarity and uniqueness.

We want to point out that the striving for uniqueness differs crucially from negative influence. Negative influence presumes that individuals want to increase existing differences from *dissimilar* others. This motivation is stronger the more *dis*similar the respective other person is. Contrary to this, the striving for uniqueness implies that individuals are trying to create differences from those they consider to be too *similar*. As soon as this similarity has decreased, the motivation to individualize will also decrease.

I.3.4. A structural factor: demographic faultlines

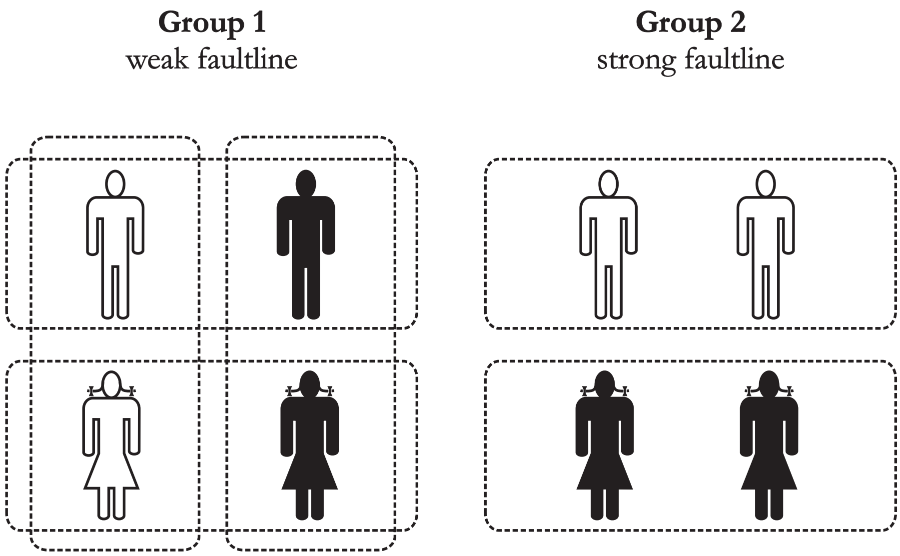
In Example 1, four passengers survived a plane crash and then try to escape from a jungle. There are two males and two females. Two are black and two are white. Classical sociological theorizing on social differentiation in modern societies would suggest that demographic differences such as these can influence the outcomes of group processes (Bourdieu 1984[1979]; Elias 1969[1939]; Simmel 1957). These theories hold that individuals want to set themselves apart from distinct social categories. To signal that they are different from distinct groups, individuals strategically reject cultural tastes, attitudes and behaviors which they consider typical for these groups or which have been adopted by outgroup members (Berger and Heath 2008; Bryson 1996).

Along the same line of reasoning, socio-psychological theories of intergroup relations propose that humans want to accentuate differences between distinct social categories (Brewer 1991; Tajfel and Turner 1986; Turner 1987). These theories hold that as an ingroup becomes psychologically salient, individuals will adjust their opinions in such a way as to conform to the ingroup's stereotypical position (Turner 1987). This position will be similar to the average opinion of the ingroup members. In contrast, when an outgroup is salient too, the stereotypical position will shift in a way so as to "distinguish the ingroup clearly from relevant other groups and thus *maximize* differences between ingroups and outgroups" (Hogg, Turner and Davidson 1990 p. 80). To summarize, theories of social differentiation and social-categorization theories hold that demographic differences can trigger negative influences. This might lead to polarized opinions in the sense that, for instance, the two black survivors will want to go in the opposite direction from the two whites.

In the group of survivors, demographic diversity is very high in the sense that the group can be separated along two dimensions, gender and skin color. Intuitively, one would expect opinion polarization to be very likely in this group. However, Lau and Murnighan (1998; 2005) proposed that the effect of demographic diversity may decisively depend on the way demographic attributes are distributed across the two dimensions (see also Colson 1954; Evans-Pritchard 1939; Flap 1988; Galtung 1966; Lijphart 1977; Ross 1920; Simmel 1922 (1908)). They claim that demographic diversity causes opinion polarization only when the distribution of demographic attributes generates a *strong faultline*. "Group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of resulting subgroups. In contrast, faultlines are weakest when attributes are not aligned and multiple subgroups can form" (Lau and Murnighan 1998: 328).

Figure I.3 illustrates Lau and Murnighan's faultline concept. The figure shows two groups of four actors each. Similar to the plane-crash survivors, both groups consist of two males and two females, on the one hand, and of two blacks and two whites, on the other. Thus, both groups are characterized by a high demographic diversity. However, the two groups differ in faultline strength. In Group 1, the two demographic dimensions are not aligned. The boxes indicate that each group member shares one demographic attribute with two other members of the group. This similarity might prevent negative-influence tendencies and therefore hamper opinion polarization. In Group 2, however, there are two white males and two black females. Thus, pairs of group members are either perfectly similar or perfectly dissimilar. The clear-cut divide between the two subgroups might cause negative influence tendencies and thus lead to opinion polarization.

Figure I.3: Two groups with identical diversity but different faultline strength



I.4. Our methodological approach: ABC modeling

In each chapter of this book, we added a new ingredient to the existing social-influence models and demonstrated how each of them helps to answers our research questions. Furthermore, in each chapter we pointed out *counter intuitive* and unexpected implications of our new models. For example, we claimed in chapter III that opinion polarization can emerge even though individuals do not seek to increase opinion differences with others. More precisely, we proposed that perfectly homogeneous populations can fall apart into subgroups which develop increasingly distant opinions even though individuals do not seek to increase opinion differences. This proposition contradicts intuition as well as standard theories of intergroup processes (Brewer 1991; Tajfel and Turner 1986; Turner 1987) and classical theories of social differentiation

(Bourdieu 1984[1979]; Elias 1969[1939]; Simmel 1957). Contrary to our approach, these theories predict that opinion differences will increase only in settings where there are already initially salient differences between subgroups of individuals. Furthermore, these theories hinge on the assumption that individuals seek to increase opinion differences with dissimilar others (negative influence). To corroborate our provocative propositions, we developed formal models of our social-influence theories. With these formal models we demonstrated that our propositions follow consistently from the assumptions we made.

In this book we developed so-called *agent-based computational models* (ABC models). In line with social-influence theories, ABC models (Bonabeau 2002; Macy and Willer 2002; Macy and Flache 2009; Smith and Conrey 2007) study interactions between multiple actors, called agents, who react to influences they receive from one another.

To develop an ABC model of a given theory, the theory's assumptions about the individuals' behavior are translated into a formal language and implemented in a computer program. Starting form given initial conditions (e.g. a certain initial opinion distribution, etc.), the behavior of each agent (which follows from the theoretical assumptions) is then calculated by the computer program. Dynamics are broken down into multiple consecutive events. For example, in ABC models of social-influence theories, the computer program typically picks an agent and updates this agent's opinion. In the subsequent event, another agent's opinion is updated. In this way, opinion changes at the previous event(s) are taken into account. In most cases, events are iterated until dynamics reach equilibrium.

The advantage of ABC models is that this method allows studying dynamics in big populations of interdependent individuals and including multiple nonlinear and probabilistic assumptions. For the classical social-influence models, central propositions have been proved using analytical methods (Abelson 1964; Berger 1981; DeGroot 1974). For instance, it has been demonstrated that opinions always converge to perfect uniformity unless no subset of actors is perfectly separated from the population. The Bounded Confidence model (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002), however, is already so complex that analytical results are available only for a limited part of the parameter space (e.g. for small groups and large and homogeneous confidence intervals) (for an overview of available analytical results see Castellano, Fortunato and Loreto 2009; Lorenz 2005; Lorenz 2007). In this book, we added further ingredients to the existing models and thereby increased their complexity even more. This suggests that analytical results are hard to obtain for the propositions presented in this book. Accordingly, we followed recent work on extensions of conventional social-influence models (see e.g. Baldassarri and Bearman 2007; Mark 2003) and employed computational agent-based modeling.

I.5. Overview of the chapters

In each chapter of this book, we discussed a potential answer to our research questions. In doing so, we always took the existing social-influence models as a baseline and added new assumptions. Thus, all models started from the assumption that individuals tend to interact with those members of their group who hold similar opinions. Furthermore, during interaction, individuals exert influence on each others' opinions.

In **Chapter II**, we added the negative-influence assumption. Recently, several publications have taken this step and have demonstrated that social-influence models can generate opinion polarization when the negative-influence assumption is included (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Mark 2003; Salzarulo 2006). We have contributed to this finding by including a second ingredient, demographic attributes. To be precise, we studied the effects of demographic faultline strength on the opinion dynamics in groups.

In Chapter II, we applied our social-influence model to work teams. Why? Because empirical research suggests that opinion dynamics in a team can affect the team's work performance. In particular, research has confirmed that consensus on opinions decreases emotional and work-related conflicts in the team (Harrison et al. 2002; Jehn 1994; Jehn, Northcraft and Neale 1999). Fewer conflicts, in turn, increase team performance and individual work satisfaction (Jehn 1994; Jehn 1995). However, work teams that perform non-routine tasks might also profit from some disagreement because this can promote inspiring discussions and critical evaluation of problems and decision options (Jehn 1995; Jehn 1997; van Knippenberg, De Dreu and Homan 2004). It has been found that there might be an "optimal level of task conflict in nonroutine-task groups" (Jehn 1995: 275). This optimum is reached when there is enough disagreement to trigger an inspiring discussion, while, at the same time, there is also enough agreement to prevent conflicts.

Furthermore, in an increasingly globalized economy, the *demographic* composition of work groups has become a central issue for organizations (Bowers, Pharmer and Salas 2000; Milliken and Martins 1996; Pelled 1996; Stewart 2006; van Knippenberg and Schippers 2007; Webber and Donahue 2001; Williams and O'Reilly 1998). Diversity in dimensions such as ethnic background, religion and gender can be beneficial for organizations, because it broadens the social and human capital that teams can use to fulfill their tasks. At the same time, demographic differences can cause disagreement and tensions which threaten performance. Out of these considerations, the very practical question emerges: Under what conditions will demographically-diverse work teams manage to overcome disagreement and be able to make use of their advantages?

Our analyses in Chapter II suggested that work teams with high demographic diversity but a weak faultline will not suffer from polarization. However, the stronger the faultline the more likely opinions will polarize. Furthermore, we demonstrated that the model highlights a new structural condition that may give managers of work teams a handle for tempering faultline effects. We argued that managers might do well to manipulate the *timing of contacts* in work teams. It follows from our model that newly formed teams with strong demographic faultlines very likely polarize when group members freely interact with each other. Because of the high demographic dissimilarity, team members reject the opinions of dissimilar colleagues and develop extreme opinions. However, we showed that opinion polarization is less likely when teams are first separated into demographically homogeneous subgroups and are merged only later in the opinion-formation process. Because of the demographic similarity between members of subgroups, it is very unlikely that they will influence each other negatively. Hence, subgroups will likely develop local consensus on moderate opinions. When subgroups are merged after they have found local consensus, opinion similarity will act as a bridge over the demographic faultline and lead to further convergence.

Besides the obvious practical implications of this intervention, the timing of contacts hypothesis also challenges a prominent theory of intergroup relations; contact theory (Allport 1954). Contact theory holds that contact between members of different social categories will improve interpersonal relations. In contrast, our model implies that contact between dissimilar group members in the early phase of the opinion formation process might intensify initial opinion differences and impair interpersonal relationships.

In **Chapter III**, we developed an alternative explanation for opinion polarization. We started by pointing out that prominent social-psychological theories of group processes (Brewer 1991; Turner 1987) hinge on the assumption that individuals seek to increase opinion differences with dissimilar others. However, empirical research on negative influence provides very little evidence for this mechanism (e.g. Krizan and Baron 2007), suggesting that negative influence shapes opinion dynamics only in a restricted number of settings. In Chapter III, we proposed an explanation for opinion polarization which does not hinge on the negative-influence mechanism. Our counter-intuitive proposition is that perfectly homogeneous populations can also split up into subgroups with maximally distinct opinions, even when individuals do not strive for any opinion differences. We developed a new social-influence model that includes the persuasive-argument mechanism and demonstrated that this model can generate opinion polarization. In addition, our analyses revealed that a *strong* individual tendency to interact with those group members who hold similar opinions is a central precondition for polarization.

Chapter III also reported the results of a laboratory experiment (N=96) which we designed to test the new theory. In this experiment, participants discussed an opinion in a computer network. We created a setting which rules out negative-influence tendencies and tested under what conditions there was opinion polarization. It turned out that there was no polarization when participants only informed each other about their opinions. However, in those experimental conditions where participants exchanged arguments, there was significant polarization. Furthermore, we included a condition where participants exchanged both arguments and opinions. In this condition, we also found significant opinion polarization. In the end, the experiment confirmed the hypothesis that the selection of similar interaction partners is an important precondition of polarization.

In **Chapter IV** we applied the model which we have developed in Chapter III to work teams. As in Chapter II, but now with the new model from Chapter III, we studied the effects of faultline strength on opinion polarization. We showed that also the argument-exchange model supports the faultline hypothesis. However, the new model points to several conditions for this effect which previous contributions have overlooked. First, even with a very strong faultline, opinions will only polarize in those groups where individuals tend to select similar interaction partners. Second, polarization is more likely, the stronger the opinions and demographic attributes in a team are correlated initially, that is, prior to interaction between the group members. Furthermore, the new model implies that the short-term effects of demographic faultlines differ crucially from their long-term effects. Groups where demographic attributes are not perfectly correlated will eventually arrive at consensus, even though they might suffer from polarization in the short run. Counter-intuitively, the model also implies that the convergence process is faster, the *stronger* the demographic faultline is. In other words, even though teams with strong faultlines might suffer from opinion polarization in the short run, they might arrive at consensus faster than teams with a weak faultline.

In **Chapter V**, we focused on the uniqueness mechanism and tested whether it helps explain opinion clustering. We took the existing social-influence models as the baseline and included the assumption that individual opinions are also shaped by the need to hold a unique opinion. In an initial step, we included striving for uniqueness as random individual opinion perturbations (white noise). However, we showed that this fails to explain opinion clustering. Weak opinion noise triggers social-influence cascades that lead to consensus. Increasing noise implies rampant individualism rather than clustering.

We presented a new solution to the clustering problem and showed that the new model does generate clustering. The key element of our model is an adaptive kind of noise. We included the factor that the individual striving for uniqueness increases when many members of the population hold similar opinions. However, when actors agree with only a few others, then they are sufficiently unique and do not seek to individualize. In a computational experiment, we identified conditions under which populations develop clusters with diversity between and consensus within clusters. Once they have developed, the clusters are temporarily stable but may later merge. Merging, however, triggers the development of new clusters. In this way, clustering is a stable outcome. Paradoxically, the new model's predictions are not only robust to noise, but noise is the central mechanism that causes cluster formation.

Table I.1 summarizes which ingredient is studied in each of the four chapters.

Table I.1: Ingredients added to the existing social-influence models

	Chapter II	Chapter III	Chapter IV	Chapter V
Individual level mechanism				
Negative influence	×			
Persuasive arguments		×	×	
Striving for uniqueness				×
Social-structural condition				
Demographic faultlines	×		×	

I.6. What have we learned and where do we go from here?

In the following, we will summarize the main strengths and weaknesses of each of our four answers to the research questions. We will also point towards elements for future research.

I.6.1. Negative influence

Recently, several researchers have integrated the negative-influence assumption into the existing social-influence framework (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Salzarulo 2006). In line with their publications, we showed in Chapter II that this approach helps to answer one of our research questions: The new model can explain opinion polarization.

The new model, however, does have two main shortcomings. First, the negative-influence model fails to generate opinion clustering. To be more precise, the model is able to explain the development of distinct clusters. However, these clusters will always hold maximally extreme opinions. This is because members of two emergent clusters will either exert positive influence on each other and find consensus, or the members will influence each other negatively and will become maximally dissimilar. In other words, the model predicts either minimal diversity (consensus) or maximal diversity (polarization). Only in a very few scenarios does the model generate opinion distributions with opinion diversity and, at the same time, some moderate agents. Such opinion distributions consist of maximally extreme clusters and moderate agents. The influences which the extremists exert on the moderate agents balance each other out and the moderate agents do not change their opinions anymore. However, this equilibrium is fragile and can be destroyed by minute opinion changes on the part of a single agent. For instance, if one of the extremists happens to develop a slightly less extreme opinion, then the moderate agents are "pulled" less strongly towards this agent. As a consequence, the influence from the opposite extreme cluster on the moderates will exceed the influence from their opponents. Hence, the moderate agents will adjust their opinions and will become extreme, too.

In sum, the negative influence model generates either minimal opinion diversity (consensus) or maximum opinion diversity (polarization). We consider it a weakness of this model that it fails to generate opinion distributions that fall between these two extremes.

The second and even more problematic shortcoming of the negative-influence model is discussed in Chapter III. Empirical research on negative influence has led to very mixed results. There are publications (Berscheid 1966; Mazen and Leventhal 1972; Sampson and Insko 1964; Schwartz and Ames 1977; van Knippenberg and Wilke 1988) which claim that they found support for negative influence. However, in Chapter III we pointed out methodological shortcomings to these studies. In addition, there are also studies which did not confirm the negative-influence assumption at all (Hogg, Turner and Davidson 1990; Krizan and Baron 2007; Lemaine 1975).

We do not claim that the negative-influence assumption should be considered falsified and excluded from social-influence models. However, we do need to be careful when we apply the negative-influence model to specific real-life settings. We cannot be confident about the model's predictions when we are not sure that negative influence might play a role in a specific setting. We therefore feel that future empirical research is needed to identify the conditions for negative-influence tendencies. For instance, individuals may be influenced negatively by dissimilar others only when they perceive significant opinion differences in the overall population. In other words, negative influence may be generated only once other factors have set opinion polarization in motion. Considering the methodological difficulties of disentangling negative from positive influence (see Chapter III), we suggest developing laboratory experiments to test hypotheses about the conditions of negative-influence.

In addition, the timing-of-contacts proposition points to an indirect test of the negative-influence assumption. In Chapter III, we demonstrated that opinion polarization is less likely when dissimilar agents are first separated and brought in contact only later in the influence process. This proposition critically hinges on the negative-influence assumption. This, in turn, implies that empirical support for the timing-of-contacts proposition indirectly confirms the hypothesis that negative influence shapes opinion dynamics. Hence, timing-of-contacts would appear to be an interesting independent variable for future experimental research.

I.6.2. Persuasive arguments

In Chapters III and IV, we showed that including the persuasive-argument principle in the social-influence framework helps to explain opinion polarization. What is more, our new model is able to generate opinion polarization even when there are no opinion differences between the agents at the outset. A central strength of the persuasive-argument mechanisms is that empirical research has consistently confirmed that argument exchange shapes opinion influence. More to the point, the exchange of arguments can intensify opinions in the sense that individuals develop more extreme views when they learn new arguments which support their opinions (Myers 1982).

However, this model also has weaknesses. First, also the persuasive-argument model is unable to explain clustering. Like the negative-influence model, opinion variance is either minimal or maximal at the end of the influence process. According to the persuasive-argument model, non-extreme clusters will not be stable because their members will interact with members of other clusters. In this way, they will likely be exposed to arguments which will change their opinion. Such arguments will then spread in their cluster and influence the remaining cluster members.

A second weakness of the persuasive-argument model is that opinion polarization is very fragile. In the current version of the model, opinion polarization is stable because members of distinct groups refuse to exchange arguments. If one includes, however, a very small interaction probability for pairs of agents that would refuse to interact otherwise, there will be argument exchange between the subgroups. As a consequence, opinions will converge in the long run. This, however, is not in line with the results from empirical research. Empirical studies on work teams in organizations suggest that, even though team members interact frequently, teams sometimes fail to overcome opinion differences (Early and Mosakowski 2000). The current model fails to explain this finding.

Nevertheless, the persuasive-argument mechanism has found considerable support in empirical research and therefore offers a promising solution to the polarization problem. Most of this empirical research, however, has been performed in the laboratory (Isenberg 1986; Myers 1982). More research outside the laboratory is needed.

Online discussion sites appear to be a promising research field. It has been argued that people tend to visit those discussion sites where they expect to meet people who hold similar opinions (Sunstein 2008). As an example, there are lively discussions about American healthcare reform on the web pages of the Tea Party movement (Tea Party Patriots 2010) and its political counterpart the Coffee Party movement (The Coffee Party USA 2010). If the supporters of a new healthcare system discuss it mainly on the Coffee Party's web pages and if its opponents mainly discuss it on the Tea Party pages, then we might observe opinion polarization during the discussions. One advantage of Internet studies is that they provide the researcher with very detailed information about the opinions of the discussants, their discussion partners and the arguments they use. Furthermore, online discussions can be tracked over very long periods. This allows one to study under what conditions polarization will be stable in the long run. This information might help develop a new persuasive-argument model which can explain stable group splits.

I.6.3. Striving for uniqueness

In Chapter V, we introduced the assumption that individuals strive for uniqueness in the sense that they want to hold opinions that are shared by relatively few others. When too many others hold similar opinions, individuals are motivated to change their opinions (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980). We demonstrated that our new model is able explain that opinion clusters emerge and remain temporarily stable. Clusters can merge but the resulting group will not be stable and will split up into clusters again. In this way, clustering is a stable phenomenon.

Our new model, however, offers a somewhat problematic solution to the polarization problem. To be more precise, our model can generate polarization. When clusters have formed, they independently perform a random walk through the opinion space. As a consequence, clusters can happen to develop increasingly distant opinions. This constitutes opinion polarization and the model thus offers a solution to the polarization problem. Still, such scenarios are *random* events. Psychological research on discussion groups, however, suggests that under certain conditions discussion groups *systematically* develop increasingly extreme opinions (Isenberg 1986). What is more, the laboratory experiment which we presented in Chapter III has found statistically significant polarization tendencies. Such systematic polarization tendencies cannot be explained with the new model.

Nevertheless, empirical research supports the assumption that individuals strive for uniqueness (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980). Future empirical research is needed to test whether this striving affects opinion dynamics in groups. For this purpose, online discussion sites appear to be a good setting. In contrast to group discussion experiments (Johnson and Johnson 1982), for example, users of online discussion sites form *large* communities. When a consensus emerges, users may thus feel that there are too many others around who hold opinions similar to the ones they do. As a consequence, people might feel less unique and change their opinion. In standard experimental discussion groups, such dynamics may be less likely because of the small size of the discussion groups (usually N=4).

I.6.4. Demographic faultlines

In Chapters II and IV we studied the effects of demographic faultlines on opinion dynamics. We showed that the negative-influence model and our model with persuasive arguments both predict that opinion polarization is more likely the stronger the demographic faultline is.

Recent empirical tests of the faultline hypothesis have led to inconsistent findings. Some studies found that strong faultline groups tend to suffer from little integration and from emotional conflicts between the team members (Early and Mosakowski 2000; Li and Hambrick 2005; Rico et al. 2007). However other studies found that faultline strength inhibits conflicts and improves psychological safety, job satisfaction and learning behavior in work teams (Gibson and Vermeulen 2003; Hart and Van Vugt 2006; Lau and Murnighan 2005). Thatcher et al. (2003) found curvilinear effects for faultline strength. They report more conflicts and lower performance in teams with weak and very strong faultlines than in teams with moderate faultline strength.

In trying to explain these inconsistent findings, researchers have included moderating variables in their statistical models. It has been shown that faultline effects increase in strength as the distance[2] between faultline subgroups increases (Bezrukova et al. 2009; Molleman 2005). High team identification was shown to prevent negative faultline effects (Bezrukova et al. 2009). Furthermore, team autonomy seems to worsen the effect of faultlines on cohesion and integration (Molleman and Slomp 2006; Rico et al. 2007).

In Chapter III, we pointed out three additional variables that might moderate the effects of faultline strength on opinion polarization and conflicts in work teams. First, the persuasive-argument model implies that strong demographic faultlines breed opinion polarization only when team members tend to interact with similar others. Second, opinions and demographic attributes already need to be correlated at the beginning of the social-influence process. Third, time may play a crucial role. In particular, the model implies that teams with a strong (but not maximally strong) faultline will overcome polarization in the long run. What is more, even though strong faultline teams might suffer from polarization in the short term, these teams might end up reaching consensus faster than weak faultline teams. This would suggest that results of empirical studies may critically depend on when opinion differences and conflicts are measured. Future empirical work should therefore focus on longitudinal data.

---

[2] For example, two 30-year-old team members differ more from two team members aged 50 than from two team members aged 40.

I.6.5. Achievements and prospects in a nutshell

In this book, we sought to explain two phenomena that existing social-influence models fail to explain; opinion clustering and opinion polarization. Our strategy was to start from the existing social-influence framework and to include new ingredients. We developed several formal models and studied whether and under which conditions each of them are able to generate clustering and polarization.

We demonstrated that each of the models provides an answer to one of our research questions. The models with negative influence and persuasive arguments are able to explain opinion polarization. The uniqueness model is able to explain opinion clustering. However, none of the three models offers a satisfactory explanation for both clustering and polarization.

Future work is needed to test the new models empirically. First, empirical research should test the assumptions we have added to the existing models. This is of central importance for the negative-influence assumption. Second, empirical research should test the new predictions of our models. Based on the results of these empirical tests, one will have to decide which of the models is most promising and is capable of forming a reliable basis for future modeling work. Alternatively, two of our ingredients could be integrated into one model.

In this book, we offered new solutions for the hitherto unresolved puzzle of opinion diversity despite social influence. Much work remains to be done, but we believe that the theories put forward in this book open up directions for fruitful future research. Moreover, they suggest potential new approaches for the management of opinion dynamics in groups that may help practitioners to avoid polarization or, if necessary, sustain a healthy diversity of opinions.

II. Negative influence and demographic faultlines[3]

Abstract

In this chapter, we will discuss the first approach to the polarization problem. We will develop a formal model that includes the assumption of negative influence and demonstrate that this model can indeed explain opinion polarization. Recently, several models that assume negative influence have been developed. We will contribute to this literature and show how demographic attributes might interfere with opinion dynamics generated by negative influence. In particular, we will discuss Lau and Murnighan's work on the effects of demographic faultlines. Lau and Murnighan proposed that polarization is the more likely the stronger demographic differences between group members correlate across various dimensions.

Computational experiments will demonstrate that the central claims of Lau and Murnighan's theory are consistent with the model. Furthermore, we will show that the model highlights a new structural condition that may give managers a handle to temper the negative effects of strong demographic faultlines. We will call this condition the *timing of contacts*. Computational analyses will reveal that negative effects of strong faultlines critically depend on *who* is *when* brought in contact with *whom* in the process of social interactions in the team. More specifically, we will demonstrate that faultlines have hardly negative effects when teams are initially split into demographically homogeneous subteams that are merged only when a local consensus has developed.

II.1. Introduction

The demographic diversity of a work team is seen as one of the major determinants of its performance. While managers as well as diversity researchers emphasize that diverse teams benefit from their large variety of social and human capital resources, (e.g. Chatman et al. 1998), many studies also highlight that this benefit comes at a potentially large cost. Diverse teams may be less socially cohesive than homogeneous teams and social cohesion, in turn, can be an important antecedent of performance (e.g. Jehn and Bezrukova 2004; Jehn, Northcraft and Neale 1999). Milliken and Martins concluded that "diversity thus appears to

[3] This chapter has been published together with Andreas Flache (first author) under the title "How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams" in Computational and Mathematical Organization Theory (2008:14/1). The article is freely available online (www.springerlink.com/content/0l03554v37118683)

be a double-edged sword" (Milliken and Martins 1996: 403), reflecting the mixed research evidence that produced both positive as well as negative effects of demographic diversity on team performance (for comprehensive reviews about theoretical and empirical research see: Bowers, Pharmer and Salas 2000; Milliken and Martins 1996; Pelled 1996; Stewart 2006; Webber and Donahue 2001; Williams and O'Reilly 1998).

The mixed effects of diversity have been attributed to the simultaneous operation of both positive effects on a team's human capital and negative effects on team cohesion (Reagans and Zuckerman 2001). However, Lau and Murnighan (1998; 2005) have questioned that demographic diversity is necessarily a threat for team cohesion. In Lau and Murnighan's view, cohesion suffers in a diverse group only to the extent that the distribution of demographic attributes across group members generates a *strong demographic faultline*. „Group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of resulting subgroups. In contrast, faultlines are weakest when attributes are not aligned and multiple subgroups can form" (Lau and Murnighan 1998: 328). To give an example, a faultline is strong in a team consisting of two Caucasian, highly educated women and two African-American men with low level of education. In this case, all three demographic dimensions along which team members differ (race, sex, educational level) split the team along the same line. The faultline would be weaker if, for example, the two highly educated team members would be one man and one woman. The core prediction (see Lau and Murnighan 1998: 331) is that stronger demographic faultlines increase the potential for dissensus between team members and thus put performance under pressure. The theory also implies that the direct effects of diversity on performance are positive due to larger human and social capital in diverse teams. Subsequent empirical research has provided partial support for the proposed negative effects of strong faultlines (e.g. Lau and Murnighan, 2005; Molleman, 2005; Thatcher, Jehn and Zanutto, 2003) and has identified organizational design features that interact with the effects of faultline strength on team outcomes, such as team empowerment strategies or the use of knowledge management systems in team learning (Gibson and Vermeulen 2003)

In a nutshell, the theory of faultlines (Lau and Murnighan 1998: 332-333) is based on two main mechanisms: First, it is assumed that team members prefer to interact with those team members who are similar with respect to a salient demographic attribute. This corresponds to the prominent notion that homophily (Lazarsfeld and Merton 1954) is a strong force in social interactions (McPherson, Smith-Lovin and Cook 2001). Which demographic attribute is salient in a certain work situation changes from situation to situation. Secondly, if actors choose to interact they are assumed to exert social influence (Festinger, Schachter and Back 1950) upon each other. Lau and Murnighan seem to assume furthermore that demographically similar actors tend to hold similar opinions even prior to interaction. Based on psychological research on opinion formation in groups (Isenberg 1986; Vinokur and Burnstein 1978) the authors propose that interactions between demographically similar actors reinforce the opinions they hold prior to interaction and, in the process, increase dissensus with demographically dissimilar group members. In other words, demographically similar interaction partners become more convinced of their respective opinions, because they tend to agree in opinion and they learn new arguments that are in line with their opinion. But only in teams with a strong faultline, the same team members interact again and again so that the opinions of the demographic subgroups become increasingly distinct at the expense of lower cohesion of the team as a whole. By contrast, in teams with weak faultlines, group members repeatedly interact with colleagues with a large variety of demographic characteristics and opinions, such that no self reinforcing dynamic towards an emergent subgroup split can develop.

While applications of faultline theory clearly demonstrate its relevance for both researchers and managers, neither Lau and Murnighan's original elaboration nor subsequent extensions have fully explicated the mechanisms that may underlie faultline effects. Both the transparency of Lau and Murnighan's theory as well as its deductive power can benefit considerably from a formal deduction of their central claims and an analysis of the precise combination of assumptions that is needed to derive them. In a previous paper we proposed a formal model of faultline effects that allows to generate hypotheses in line with previous informal reasoning (Flache and Mäs 2008b). We could also show that it is not even necessary to assume that opinions and demographic characteristics of team members are correlated already prior to interaction[4]. In the present paper, we move one step further and argue that the model also implies a remedy against negative effects of strong faultlines that has hitherto been overlooked in the literature. We propose that the effects of strong faultlines may critically depend on *who* is *when* brought in contact with *whom* in the process of social interactions in the team. More generally, it may depend on the *timing of contacts* between team members whether strong faultlines have negative effects on team cohesion. To be precise, we use "timing of contacts" here in the

[4] We explain this point below in our elaboration and discussion of the formal model.

sense of Moody's (2002) concept of "relationship timing". Broadly, relationship timing defines the sequence within which social interactions occur in given network of interactions. Consider for example an opinion formation process between three members of a work team, two of whom agree with each other and totally disagree with the third one. One possible timing of contacts might be that all three group members are brought together to discuss the issue. In this case, social influence occurs simultaneously in all three dyads in the network. Another sequence might be that only one of the two majority members discusses the issue with the minority member and after each meeting, the two majority members come together again. Obviously, in the first sequence the deviant might influence the positions of both other team members at the same time, while in the second sequence, he can directly influence only one of them, while the other one may bring his colleague "back into line" after each encounter with the deviant.

Effects of the timing of contacts on the outcome of group discussions have been demonstrated in experimental research by Kameda and Sugimori (1995). These authors manipulated the sequence within which in a group discussion minority members encountered majority members and found that this affected the chances for consensus in the overall group. More recently theoretical analyses have shown that the diffusion dynamics of, e.g., knowledge or diseases in social networks may critically depend upon the timing of network contacts (Gibson 2005; Moody 2002). For example, whether an infectious disease can spread in a chain A-B-C from A to C critically depends upon whether B was infected by A before or after being in contact with C. The idea that timing matters has not yet been theoretically elaborated for the study of opinion dynamics. However, we believe that the diffusion of opinions may be similarly affected by relationship timing than the diffusion of infection or information. The key reason why we expect the timing of contacts to be important for the group dynamics in diverse teams is the inherent path dependence of the process of social interactions between team members. For example, early contacts between group members who are strongly dissimilar both in terms of their opinions and their demographic characteristics may trigger negative and hostile interactions between the interactants. This, in turn, may lead them to adopt extreme positions on some issues. If these "radicalized" actors interact subsequently with demographically similar "friends", this may entail "bandwagon dynamics" in which the friends of the early conflict partners are socially influenced to adopt similarly extreme positions. The stronger are demographic faultlines, the more such a dynamic would project the demographic faultline onto an emergent faultline in the opinion space, with the result that communication between team members and thus group cohesion and team performance may severely suffer. Clearly, this downward spiral might be avoided when contacts between team members are arranged in such a way that opposed "extremists" are initially isolated from each other and are instead exposed to interactions with demographically similar in-group members who are more moderate in their opinions. Then, the likely consequence is that initial extremists also become more moderate and initial moderates from different demographic subgroups move towards each other in the opinion space.

It may be a plausible idea that the timing of contacts modifies faultline effects, but Lau and Murnighan's original theory is not precise enough to generate testable predictions about the exact conditions under which this mechanism may work. We use and extend in the present analysis the formal model proposed by Flache and Mäs (2008b) to elaborate our reasoning why timing matters and under which conditions. In section 2, we describe the formal model and its extension to accommodate timing effects. Section 3 contains a description of the simulation experiments and results. In section 4, we discuss results and offer conclusions.

II.2. The Model

The model consists of four main elements, the formalization of the dynamics and elementary mechanisms of *social interactions and influence* between team members, the operationalization of *demographic faultlines*, the model of the *timing of contacts* and, finally, *aggregate outcome measures* that capture the dependent variables we are interested in.

II.2.1. The Social Interaction and Influence Dynamics

The main endogenous outcome variable of our model is the *distribution of work related opinions* in the team, because following previous work (Mason 2006: 234; Molleman 2005: 175-176; Pfeffer 1985) we assume that consensus at least on fundamental issues seems a necessary precondition for effective teamwork, while opinion polarization on these issues may be a major obstacle to good team performance. The theoretical assumptions of homophily and social influence identify a clear causal link between team cohesion, consensus on work related opinions and the strength of demographic faultlines in a team. Broadly, the stronger are faultlines in the team, the less likely it is that team members in different subgroups influence each other sufficiently to generate a consensus on work related opinions on the level of the team as a whole, and the more likely it is that the influence processes result in polarization rather than consensus. At the same time, the combined assumptions of homophily and influence link the degree of consensus closely to the level of cohesion in the team. We assume that only when team members agree on important issues, they have good social relations with each other which, in turn, generates social cohesion.

With this approach, we deliberately exclude from our analysis variables which also may affect performance but which are not or at least much less directly causally related to faultline strength (Lau and Murnighan 1998), like the size of the team's pool of human and social capital.

We assume that the effects of faultlines on opinion polarization (and poor team performance) are generated by the interplay of the four fundamental social mechanisms *homophily, social influence, heterophobia and rejection.* According to homophily[5], the more similar two actors are with respect to salient opinions or demographic characteristics, the more they like each other and the more they interact (Brass et al. 2004; Byrne 1971; Harrison and Carroll 2002; Homans 1951; Kandel 1978; Lazarsfeld and Merton 1954; McPherson, Smith-Lovin and Cook 2001; Rogers and Bhowmik 1970). According to social influence, if two persons interact they adapt their opinions (Abelson 1964; Brass et al. 2004; Kerr and Tindale 2004). But homophily and social influence alone do not suffice to explain why groups with strong faultlines exhibit a tendency towards extreme and over time increasing opinion differences between a small number of opposed and demographically dissimilar factions in the team (cf. Early and Mosakowski 2000)[6]. To address this pattern with our model, we followed previous research and complemented the mechanisms of homophily and social influence with their negative counterparts of *heterophobia* and *rejection* (Flache and Macy 2006b; Jager and Amblard 2005; Kitts 2006; Macy et al. 2003; Rainio 1961a; 1961b; 1962; 1965; Salzarulo 2006). Heterophobia implies that if the dissimilarity of two actors exceeds a certain threshold then the actors do not like each other (Byrne, Clore and

---

[5] Note that the definition of 'homophily' which we use in this chapter does not correspond to the definition used in the chapters I and II. In this chapter, homophily refers to the *preference* of individuals to interact with similar others. In the previous chapters, however, it referred to the *tendency* to interact with similar others. Such a tendency can result from the preference to do so, but also from social influence during interaction (see chapter II).

[6] Axelrod (1997) proposed to add the assumption that social influence may be entirely cut off when actors disagree beyond a certain critical level. With this assumption, homophily and social influence can stabilize differences between subgroups (Axelrod 1997; Flache and Macy 2006b; Flache, Macy and Takács 2006; Hegselmann and Krause 2002; Weisbuch, Deffuant and Amblard 2005). However, this explanation is not readily applicable to work groups, where there is little room to entirely avoid social interaction with dissimilar others. Moreover, Axelrod's assumptions can at best explain why differences between subgroups persist over time, but not why groups may increasingly polarize in the course of team interaction, as described by Earley and Mosakovski (2000).

Smeaton 1986; Chen and Kenrick 2002; Pilkington and Lydon 1997; Rosenbaum 1986a; Rosenbaum 1986b; Smeaton, Byrne and Murnen 1989). Rejection states that actors have a tendency to change their attributes in a way to become more dissimilar to interaction partners they do not like (Abelson 1964; Kitts 2006; Salzarulo 2006; Tsuji 2002).

It is important to note that Lau and Murnighan do not directly assume rejection. They propose instead that increasing opinion differences between dissimilar actors result from a self-reinforcing dynamic that is triggered by an initial correlation between demographic attributes and opinions. We avoided this assumption, because what we aim to explain is that the strength of the demographic faultline leads to opinion polarization along this faultline. If we already assume in the model that demographic attributes are correlated with the opinions then it is not surprising that the model predicts exactly this as an outcome. In our previous work (Flache and Mäs 2008b), we could show that the model sketched here suffices to reconstruct the main regularities predicted by faultline theory. Hence, we argue that the assumption of an initial positive correlation between demographic attributes and opinions should be avoided in this context. However, our argument is purely theoretical. We do not claim that the dynamics that Lau and Murnighan describe do not occur in real work teams.

Finally, our model distinguishes between two types of attributes on which agents can differ and which define the level of similarity between agents. Demographic attributes on the one hand are fixed and can not be changed by the dynamics of social influence and rejection. On the other hand, opinions are flexible and are subject to social influence and rejection. Previous computational studies based on similar sets of assumptions have already demonstrated how demographic differences can lead to the emergence of cultural niches in demographic space such that demographically dissimilar actors also hold dissimilar or even radically opposing opinions (Macy et al. 2003; Mark 2003). However, these studies did not address the effects of faultline strength in the demographic distribution.

Technically, each of the N team members is represented as an agent i characterized by D *fixed* (a_{id}^{fix}) and K *flexible attributes* (a_{ik}^{flex}), where d and k refer to the d'th and k'th fixed and flexible attribute, respectively. The fixed attributes correspond to the demographic characteristics, the flexible ones represent the agent's work related opinions. For simplicity, we assume that demographic attributes and opinions are equally salient. Moreover, we focus on clearly distinguishable demographic attributes, expressed by the assumption that demographic attributes are dichotomous and can take either the value -1 or +1

($a_{id}^{fix} \in \{-1;1\}$). Opinions of the team members can instead vary continuously between -1 and +1 ($-1 \leq a_{ik}^{flex} \leq +1$).

A key assumption of the model is that the direction and strength of influence that an agent *i* imposes on an agent *j* does not depend directly on the opinion of *j*, but it is moderated by the sign and the strength of the interpersonal relation between *i* and *j*. To model the *interpersonal relations* between the team members we assume a directed graph where w_{ij} represents the weight of the corresponding relationship ($-1 \leq w_{ij} \leq +1$). If team member *i* has contact to team member *j* then the weight w_{ij} takes a nonzero value between -1 and 1. A positive weight reflects that *i* evaluates *j* positively, whereas a negative one represents a hostile relationship. If there is no contact between *i* and *j*, or *i* is indifferent between liking and disliking *j*, then the weight is 0.

Both the *K* flexible attributes and the weights of the relationships are endogenous and change in discrete time steps. In every time step, one team member is selected randomly with equal probability to update either his flexible attributes or weights. With probability 0.5, all weights of *i* are updated simultaneously. In the event that flexible attributes are updated, all flexible attributes are updated simultaneously.

Time is modeled in discrete steps. The duration of a simulation run is expressed in number of iterations. One iteration corresponds to *N* simulation steps to assure that on average each agent updates either his weights or his attributes once within an iteration. To be sure, given the asynchronous random updating of agents, an iteration does not encompass any particular length of time or synchronization of events (e.g. work days). Rather, one iteration indicates that *N* events have taken place in which agents have changed their opinions or weights.

Similar to previous models of social influence with continuous opinions (Abelson 1964; Hegselmann and Krause 2002), we assume that the change of team member *i*'s flexible attribute *k* is an aggregated result of the influences imposed by all other agents who exert influence upon *i*. Technically, the new value of the attribute, $a_{ik,t+1}^{flex}$ is obtained by adding to the old value a weighted sum of the pressures of all influential others. To model a somewhat gradual change of opinions, we also divide this weighted sum by 2. The pressure imposed by a single alter *j* "pulls" *i* towards *j*'s opinion if the weight w_{ij} is positive, and "pushes" *i* away from *j*'s opinion if the weight is negative. The magnitude of this pressure is proportional to the distance in opinions between *i* and *j*, $a_{ik}^{flex} - a_{jk}^{flex}$. With only positive weights summing to one, this assumption would imply that the net pressure imposed on *i* moves the agent towards the weighted average of the opinions of all interactions partners. Equation 1 formalizes these assumptions.

$$a_{ik,t+1}^{flex} = a_{ik,t}^{flex} + \frac{1}{2C_t}\sum_{i\neq j} w_{ij}(a_{jk,t}^{flex} - a_{ik,t}^{flex}) \qquad (1)$$

The C_t in equation (1) refers to the number of agents who are in contact with *i* at the moment influence takes place ($C_t \leq (N$-1)). We will discuss further below effects of interaction structures in which team members can interact only with a subset of other team members temporarily. To be precise, equation (1) only shows the principle model of influence. In the actual implementation, we apply a slight modification of the influence equation both to make sure that opinions do not go out of bounds and to smoothen the change of opinions when agents move towards the extreme ends of the opinion scale. Equations 1a and 1b fully specify opinion change.

$$\Delta a_{ik,t}^{flex} = \frac{1}{2C_t}\sum_{j\neq i} w_{ij}(a_{jk,t}^{flex} - a_{ik,t}^{flex}) \qquad (1a)$$

$$a_{ik,t+1}^{flex} = \begin{cases} a_{ik,t}^{flex} + \Delta a_{ik,t}^{flex}(1 - a_{ik,t}^{flex}), \text{if } a_{ik,t}^{flex} > 0 \\ a_{ik,t}^{flex} + \Delta a_{ik,t}^{flex}(1 + a_{ik,t}^{flex}), \text{if } a_{ik,t}^{flex} \leq 0 \end{cases} \qquad (1b)$$

The second key element of our model is the update of weights. Following previous work (Macy et al. 2003) we assume that the weight that agent *i* has towards an agent *j*, changes depending on the similarity between *i* and *j* in terms of both their demographic attributes and their opinions. More precisely, we assume that after updating, the weight adopts a level that is proportional to the current level of similarity. The new weight is negative if the average distance between *i* and *j* across all dimensions of demographic and opinion space exceeds one, i.e. half of the maximum average distance. If this average distance is exactly one, the weight is zero and otherwise it obtains a positive value. Technically,

$$w_{ij,t+1} = 1 - \frac{\sum_{d=1}^{D}\left|a_{id,t}^{fix} - a_{jd,t}^{fix}\right| + \sum_{k=1}^{K}\left|a_{ik,t}^{flex} - a_{jk,t}^{flex}\right|}{D+K} \qquad (2)$$

II.2.2. Faultline strength

To disentangle the effects of the strength of demographic faultlines from effects of demographic diversity, we devised a method that allows for varying faultline strength and keeping diversity constant at the same time. More precisely, we generated different distributions of the fixed attributes in such a way that all fixed attributes were equally frequent (= all distributions generate equally diverse groups) but the correlation between the attributes differed between distributions (= the strength of the faultline differs).

Table II.1 shows our construction method for the prototypical case of a group with 20 members (N=20) who differ along three demographic dimensions (e.g. male / female, young / old, western / non-western ethnic background). Column 2 of the table shows that we constructed the first demographic variable (A_1) by alternately assigning the values -1 and 1 to the first $N/2$ agents beginning with the value 1 for agent 1. We did the same with the second $N/2$ agents but here we started with the value -1. The distribution of this variable is the same in all work teams.

We expressed the faultline strength by a parameter f that varies between 0.5 and 1, where f=0.5 corresponds to a situation where the demographic attributes are completely uncorrelated and f=1 imposes a perfect correlation between all demographic attributes. The first step in the construction is to impose the correlation between attribute A_1 and A_2 that corresponds to the given parameter value of f. To arrive at the values for attribute A_2, we assigned to the first ($100 \cdot f$) % of the cases the same value as for attribute A_1. For example, for f=0.9, the first 90% of the agents (= the first 18 agents if N=20) hold the same value at attribute A_1 and A_2 (See the grey cells in column 3 of table II.1). To the rest of the agents we assigned on attribute A_2 the opposite value of what we assigned for attribute A_1.

To determine the values for attribute A_3 we used the same method with a small change. We first assigned to the first ($50 \cdot f$) % of the cases the same value as for attribute A_1. Then we continued with the ($N/2+1$)th case and again assigned to the following ($50 \cdot f$) % of the cases the same value as for attribute A_1. Again the rest of the cases got the opposite value than for attribute A_1. Thus for f=0.9 and N=20 the agents 1-9 and 11-19 hold the same value at attribute A_1 and A_3 (see column 4 of table II.1). This procedure makes sure, that the agents also hold at the attributes A_2 and A_3 in exactly ($100 \cdot f$) % of all cases the same value.

Table II.1: Implementation of faultline strength

	f = 0.9			*f* = 0.8			*f* = 0.7			*f* = 0.6			*f* = 0.5		
i	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
7	1	1	1	1	1	1	1	1	1	1	1	-1	1	1	-1
8	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1
9	1	1	1	1	1	-1	1	1	-1	1	1	-1	1	1	-1
10	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1
11	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1
13	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1
14	1	1	1	1	1	1	1	1	1	1	-1	1	1	-1	1
15	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1	-1
16	1	1	1	1	1	1	1	-1	1	1	-1	1	1	-1	-1
17	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1	1
18	1	1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
19	-1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	1	1
20	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1
Σ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Correlations (*r*)

	A_2	A_3
A_1	.8	.8
A_2		.8

	A_2	A_3
A_1	.6	.6
A_2		.6

	A_2	A_3
A_1	.4	.4
A_2		.4

	A_2	A_3
A_1	.2	.2
A_2		.2

	A_2	A_3
A_1	.0	.0
A_2		.0

Table II.1 also reports the correlations between the three attributes. Note that for a given distribution all pairwise correlations between two of the three attributes are equal. The relationship between *f* and the correlation is: $r = -1 + 2f$. If *f* takes the value 1 then the three attributes are perfectly correlated (r=1) and the faultline strength is maximal. If *f* takes the value 0.5 then there is no relationship between the attributes (r=0). Thus the faultline has a minimal strength. Because of its better intelligibility we use the parameter *r* to describe the faultline strength in the following. At all levels of *f*, all variables are equally distributed in all teams.

The key advantage of our method is that it separates variation in faultline strength from variation in diversity. A more intuitive alternative approach could have been to assign attributes randomly with a given probability and a given correlation. However, for the relatively small groups we are interested in, that method would have produced considerable random variation in faultline strength between single realizations of distributions imposed by the same level of *f*. Our deterministic approach excludes this source of random noise and thus allows us to focus in our computational experiments exclusively on effects of variation in *f*.

II.2.3. Timing of contacts

There is in principle an infinite number of ways how the sequence of who is when brought into contact with whom in a work team can be manipulated. For the sake of simplicity, we decided to focus upon *three ideal typical forms of timing* that we expect to shape the effects of faultline strengths in clearly different ways. The *first* form is the baseline scenario: *no timing*. Under "no timing" there is from t=1 on no structural restriction on the interactions between group members. I.e. all dyads are simultaneously "active" in the process of social influence, Technically, every agent can have a weight different from zero towards all other group members (C_t=(N-1), see equation 1). The *second* scenario represents the intuition that it may temper the effects of strong faultlines if in the early phases of the group process interactions are restricted to relatively homogeneous smaller subgroups (C_t<(N-1)). In other words, dyads between group members who are strongly different demographically are not active in the first phase of influence process, while dyads between similar group members are. For the sake of idealization, we represent the subgroups as isolated "caves" such that interaction in the early phase is entirely restricted within caves. All weights between agents who do not belong to the same cave are set to zero and kept at zero until caves are merged. The corresponding timing scenario of "first homogeneous caves, then complete" imposes from t=1 on isolated caves which are arranged in such a way that the demographic homogeneity within the caves is very high[7]. Then, after a critical time point t^*, the boundaries between caves are eliminated. Technically, we set at t^* all weights between agents who belong to different caves to the value that corresponds to their current overall similarity (see equation 2) and leave all others weights and opinions unchanged. Hence, after t^* agents will be influenced by all other agents in the team and can have non-zero ties with all other group members. The *third* and final scenario is a control condition that we called "first heterogeneous caves, then complete". We wanted to know whether the formation of smaller subgroups in the initial phase also tempers faultline effects when the subgroups are not homogeneous but instead are formed randomly. Our intuition is that this scenario will not differ from the baseline, because particularly in groups with strong

[7] Details of the method for maximization of demographic homogeneity depend on the exact settings for group size and cave size and will thus be explained in the discussion of the initialization of our computational experiments (section III.3).

faultlines it is likely that demographic divisions will induce early splits and extremist opinions within each separate cave. Once this local polarization arises, the merger of caves into one large group is likely to "export" extremism and thus polarization also into the group as a whole.

II.2.4. Aggregate outcome measures

The main claims of the theory of faultlines address two relationships: First the relationship between faultline strength and the level of consensus in the team, and second the relationship between faultline strength and the degree to which divisions in opinions are associated with demographical divisions in the team. To assess whether our model can reproduce these relationships, we devise four different outcome measures, *opinion diversity* , *opinion variance, polarization* and a measure of the degree to which differences in fixed (demographic) and flexible (opinion) attributes of agents are associated with each other, called attribute-opinion covariance, *cov(fix;flex)*.

Opinion diversity is based on a count of the number of different opinion vectors present in the group as a whole, where only flexible attributes are taken into account. For normalization, we divide this number by the group size *N*. We set *opinion diversity* =0 if there is perfect consensus. Hence, 0 ≤ *opinion diversity* ≤ 1. Clearly, both a group with high consensus and a group with perfect polarization will exhibit low *opinion diversity*. Perfect consensus implies that all agents share the same vector of opinions (*opinion diversity* = zero), whereas perfect polarization implies that there are exactly two maximally different factions in the opinion space (*opinion diversity* = 2/*N*).

Opinion variance is the average standard deviation of opinions across all *K* dimensions of the opinion space. In the case of perfect consensus, we obtain *opinion variance* = 0, and in the case of perfect polarization with two equally large maximally opposed subgroups we measure *opinion variance* = 1, the highest value we ever obtained[8]. However, a high level of *opinion variance* does not necessarily indicate that the group polarizes in the opinion space. High *opinion variance* may occur if agents strongly differ from each other in all dimensions of the opinion space, but these differences are not correlated across dimensions. In that case, the group is not polarized.

[8] In this case, the average opinion in all dimensions is zero. Moreover, in all dimensions half of the group adopts an extreme opinion at +1 and the other half of the group does so at -1. Hence, on average the distance from the mean amounts to +1 in all dimensions, yielding the result of *variance* = +1.

Polarization captures the degree to which the group can be separated into a small set of factions who are mutually antagonistic in the opinion space and have maximal internal agreement. To compute *polarization*, we use the variance of pairwise agreement across all pairs of agents in the population, where agreement is ranging between -1 (total disagreement) and +1 (full agreement), measured as one minus the average distance of opinions (averaged across all K subdimensions). This measure obviously adopts its lowest level of zero for the case of perfect consensus. The maximum level of polarization (*polarization*=1) is obtained when the population is equally divided between the opposite ends of the opinion scale at -1 and +1 and all opinion dimensions are perfectly correlated[9]. With uniformly distributed opinions, the polarization measure yields approximately 0.22 for K=1.

To test the relationship between demographic differences and differences in opinions, we compute the attribute-opinion covariance, *cov(fix;flex)* as the covariance between the vector of pairwise demographic dissimilarities and the pairwise opinion dissimilarities, where we computed for every pair of actors i and j the dissimilarity measures $\Delta_{i,j}^{fix}$ and $\Delta_{i,j}^{flex}$, as given by equations (4a) and (4b). These dissimilarity measures express the average distance across all dimensions for fixed attributes and flexible opinions, respectively. The resulting covariance *cov(fix,flex)* adopts values between -1 and 1. A value of zero indicates that similarity in opinions and similarity in demographic attributes are statistically unrelated. The initial values of *cov(fix,flex)* are expected to be near to zero, because opinions are initialized randomly. Changes of *cov(fix,flex)* that occur when the simulation proceeds indicate how much differences in opinions and demographic differences become aligned.

$$\Delta_{i,j}^{fix} = \frac{1}{D}\sum_{d=1}^{D}\left|a_{id}^{fix} - a_{jd}^{fix}\right| \tag{4a}$$

$$\Delta_{i,j}^{flex} = \frac{1}{K}\sum_{k=1}^{K}\left|a_{ik}^{flex} - a_{jk}^{flex}\right| \tag{4b}$$

Thus *cov(fix;flex)* is calculated as given by equation (5).

[9] To see this: In 50% of all dyads the agreement is 1 (indicating maximal agreement), in 50% it is -1 (indicating maximal disagreement). The average level of agreement is zero and the average distance between the agreement in a particular dyad and the average level of agreement, i.e. the variance, yields polarization = 1.

$$cov(fix, flex) = \frac{\sum_{j \neq i}\left(\left(\Delta_{ij}^{fix} - \overline{\Delta^{fix}}\right)\left(\Delta_{ij}^{flex} - \overline{\Delta^{flex}}\right)\right)}{N(N-1)} \quad (5)$$

II.3. Results of the computational experiments

We structured our computational analysis in two sets of experiments. In the *first set of experiments* the objective is to show that the dynamics of our model are consistent with Lau and Murnighan's (1998) informal reasoning. More precisely, we devise a fixed work team scenario and conduct ceteris paribus replications of the group dynamics that our model generates for different levels of faultline strength under the given scenario. The stylized regularity our model should produce in this set of experiments is a clear-cut negative relationship between the average level of consensus in the opinion distribution and the strength of demographic faultlines, *r*. More in particular, the model should generate both less often consensus and more often polarization as *r* increases. A second regularity that follows from the theory of faultline effects is an increasing association of opinion divisions with demographic divisions as faultlines become stronger. In other words, the stronger are the demographic faultlines, the clearer we expect subgroup splits in the simulated opinion distribution to reflect the distribution of demographic attributes.

The *second set of experiments* focuses on the effects of timing. Broadly, we expect that the negative effects of strong faultlines will be considerably tempered when homogeneous and mutually isolated subgroups are formed in a first phase, before in a second phase all group members interact with each other. We also want to test whether – as we intuit - this form of timing reduces the association between demographic and opinion differences in the team. To test these intuitions, we will conduct ceteris paribus replications of the scenario analyzed in the first set of experiments, but now with variation of the timing of contacts across the two options of "first homogeneous caves, then complete" and, "first heterogeneous caves, then complete", where the results of experiment 1 serve as the "no timing" baseline.

In both sets of experiments we use the following parameter settings. With regard to group size, we assume *N*=20, a size that is not too big to be unrealistic for a work team, but also large enough to allow for a sufficiently fine-grained variation in the strength of demographic faultlines (cf. Table 1). Furthermore, we assume that there are three salient demographic (fixed) attributes (*D*=3). As table 1 shows, the combination of 20 agents and

3 fixed attributes allows sufficient variation in the correlations between the fixed attributes of team members. Values for the demographic attributes are assigned to agents as shown in table 1, imposed by the data set we generated for the corresponding level of faultline strength *f*. For the number of flexible attributes (opinions), we choose K=4. This is the smallest number that makes polarization under strong faultlines not trivial, because with K=3 and D=4 it is still possible that two agents who maximally differ in all three demographic dimensions can have a positive relationship if they have sufficiently similar opinions. At the same time, this setting makes it hard to avoid polarization in a group with maximal faultline strength. Accordingly, K=3 and D=4 provides a particularly hard test for our conjecture that the right form of timing can prevent polarization even in groups with strong faultlines. Furthermore we assumed that initially (at the outset of t=1) all opinions of all agents are randomly drawn from a uniform distribution with full coverage of the entire opinion interval and with statistically independent dimensions of the opinion space. As a consequence, initial opinions are also statistically independent from demographic attributes. After initial opinions have been assigned, initial weights are computed on basis of overall similarity (see equation 2). In the timing experiments, initial weights between agents who do not belong to the same cave are set to zero and kept at zero until the boundaries between caves are removed.

II.3.1. Experiment 1: The effects of faultline strength

To illustrate how variation in faultline strength affects the model dynamics, we show first two typical simulation runs obtained for a setting with low faultline strength (r=0.2) and high faultline strength (r=0.8), respectively. Figure II.1 charts for both settings the dynamics of the four outcome measures for the first 120 iterations.

Figure II.1 shows dramatically different outcomes for the two different levels of faultline strength. In the weak faultline case, the simulated group quickly moves towards perfect consensus, as indicated by the rapid decline of *opinion diversity* and *opinion variance*, as well as *polarization*, from the levels given by the initial random distribution down to the theoretical minimum level of zero for all three outcome measures. The graph also shows that from the outset there is no (actually even a slightly negative) association between differences in opinions and demographic differences (see *cov(fix;flex)*). In the strong faultline case, it takes about 60 iterations until the group has moved from the random initial opinion distribution towards perfect polarization into two maximally opposed factions. Moreover, opinion divisions and demographic divisions align almost perfectly in this case, as indicated by a level of *cov(fix;flex)* = 0.8 obtained after about 60 iterations.

Figure II.1: Change in outcome measure for typical simulation runs with weak faultline (left) and strong faultline (right). N=20, D=3, K=4. No timing of contacts.

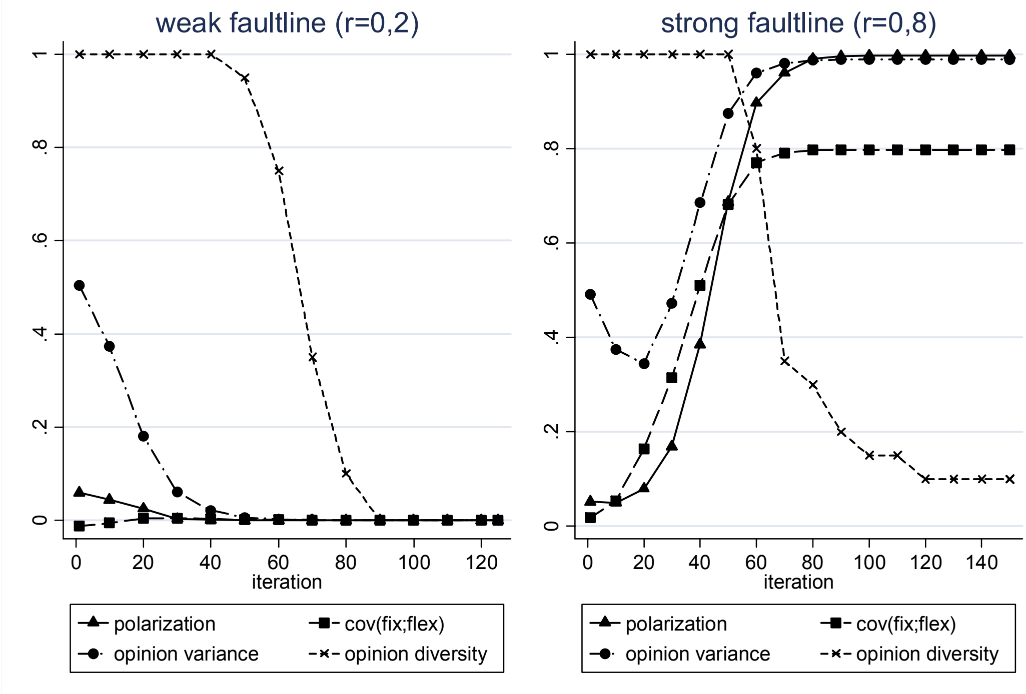


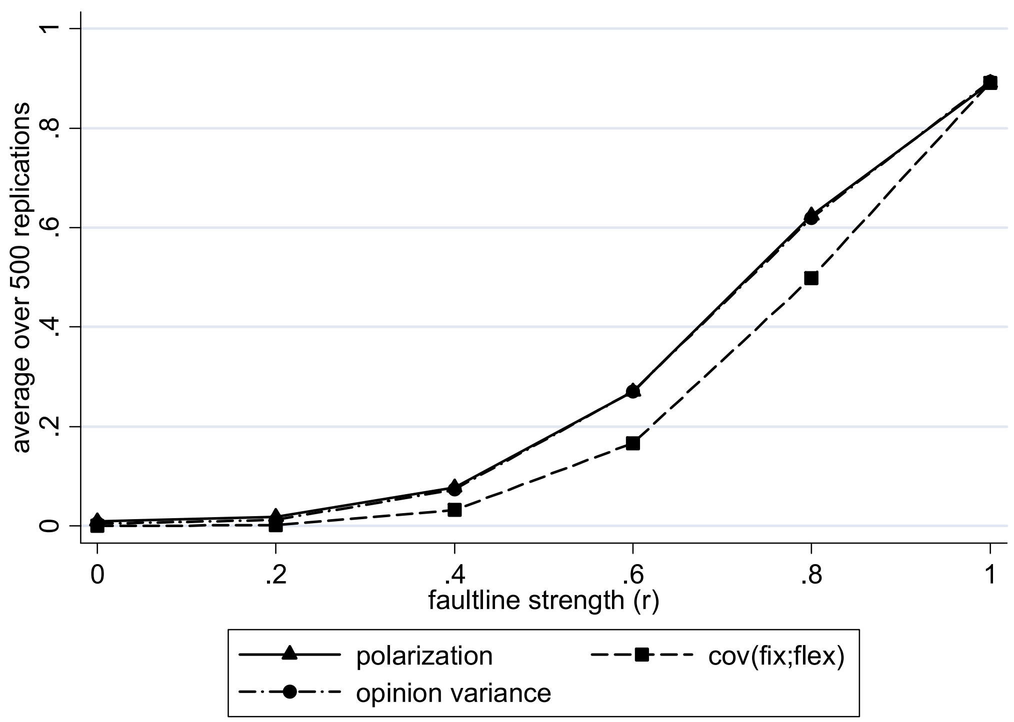
Figure II.1 shows dramatically different outcomes for the two different levels of faultline strength. In the weak faultline case, the simulated group quickly moves towards perfect consensus, as indicated by the rapid decline of *opinion diversity* and *opinion variance*, as well as *polarization*, from the levels given by the initial random distribution down to the theoretical minimum level of zero for all three outcome measures. The graph also shows that from the outset there is no (actually even a slightly negative) association between differences in opinions and demographic differences (see *cov(fix;flex)*). In the strong faultline case, it takes about 60 iterations until the group has moved from the random initial opinion distribution towards perfect polarization into two maximally opposed factions. Moreover, opinion divisions and demographic divisions align almost perfectly in this case, as indicated by a level of *cov(fix;flex)* = 0.8 obtained after about 60 iterations.

The explanation for the differences shown by Figure II.1 can be readily derived from our model assumptions. In the weak faultline scenario, demographic attributes are almost perfectly uncorrelated with each other. Hence, there are only very few pairs of agents who maximally differ on all three demographic dimensions. This makes it unlikely that negative ties (w_{ij}<0) arise in the initial configuration. In addition, if some negative ties arise, then they will most likely be between agents who are, in turn, embedded into a large number of positive ties with the same colleagues. As a consequence, positive social influence prevails and rejection hardly ever occurs in the social interactions between agents. If some agents are "pushed" to reject some enemies' opinions, then they are at the same time "pulled in" by many more friends so that the net change of their opinion is more likely towards the group average than towards the extreme ends of the opinion scale. A similar reasoning explains why the outcome for the strong faultline case is so different. In the strong faultline case, demographic differences are maximal within a large fraction of the dyads in the team. In these dyads only relatively small opinion differences in the initial configuration suffice to generate a negative relationship between the interactants. Moreover, these negative relationships tend to segregate the two major subgroups in demographic space so that most agents have the same enemies than their friends have. This entails a quick self reinforcing dynamic towards opinion polarization. Most agents move towards whatever is the current average opinion profile in their (demographic) in-group and they distance themselves from whatever is the current average opinion profile in the (demographic) out-group. The result is a coordinated movement of all agents that soon leads to convergence of their opinions on two opposite poles that align with the demographic faultline in the group.

For statistical reliability, we conducted a large number of replications of this simulation experiment and varied faultline strength across the entire interval between r=0 and r=1.0 in steps of 0.2. Figure II.2 reports the average of the outcome measures we obtained after iteration 1000, over 500 replications per condition. We do not report *opinion diversity* in Figure II.2, because final states are almost always either perfectly polarized or exhibit perfect consensus so that the variation of *opinion diversity* across conditions is extremely small. To make it easier to distinguish the different outcome measures in the Figure, we used lines to connect the data points for the six different levels of r that we simulated but we did of course not obtain results for r-values other than those shown in table II.1.

Figure II.2 clearly confirms that our model generates the stylized regularities predicted by Lau and Murnighan's theory of faultlines. All three outcome measures consistently increase with higher levels of faultline strength. More specifically, the average outcomes of almost zero for *opinion variance*, *polarization* and *cov(fix;flex)* when demographic dimensions are entirely unrelated (r=0) indicate that virtually all simulated groups have reached almost perfect consensus in this condition. By contrast, with maximal faultline strength (r=1.0) groups almost always polarize maximally, as indicated by an average *polarization* and an average *opinion variance* at the same level. The correspondingly high value of *cov(fix;flex)* in this condition shows that it is the demographic faultline along which the group also splits in the opinion space. The consistent increase of the outcome in between these two extremes shows that - for the given set of conditions (N=20, D=3, K=4) - our model clearly implies that higher faultline strength is associated with less consensus, more polarization and a stronger association between demographic and attitudinal differences, as predicted by Lau and Murnighan's theory. A further striking feature of Figure II.2 is that average polarization and opinion variance take almost the same values for all conditions. The reason for this is explained in more detail in Flache and Mäs (2008b). It is shown there that the model tends to generate in almost every replication of the experiment either nearly perfect polarization or nearly perfect consensus. The effects of faultline strength reported in Figure II.2 mainly reflect a shift in the distribution of these two outcomes. Accordingly, in a single run polarization and opinion variance take in equilibrium almost always either both the value of zero (consensus and no polarization) or of +1 (maximal variance and maximal polarization).

Figure II.2: Effect of faultline strength on outcome measures, averages over 500 replications per conditions, outcomes measured after 1000 iterations per replication N=20, D=3, K=4. No timing of contacts.



II.3.2. Experiment 2: Effects of timing of contacts

The design of our second experiment mirrors that of experiment 1, except that we replicate all conditions for the two different forms of timing, "first homogeneous caves, then complete" and "first heterogeneous caves, then complete". For the conditions that impose temporary caves, we set the size of caves to $N_C = 5$. This cave size is chosen because with N=20, it allows to easily generate demographically homogeneous caves. With $N_C = 5$ and the 50/50 distribution of demographical attributes that we use in all demographic dimensions, it is always possible to collect within one cave those 25% of the agents in the group who are equal on at least the first two of their three demographic attributes. For the condition "first homogeneous caves, then complete", we generate the corresponding caves as follows. In a first step, we lexicographically order the set of agents based on their three fixed attributes. Thus, the first five agents in this ordered set have attributes -1,-1 on dimensions d=1 and d=2 respectively, the subsequent five agents have attributes -1,+1 and so forth. In the second step, we match these relatively homogeneous subgroups of five generated in step 1 with the caves of size five.

Table II.2: Initialization of homogeneous caves

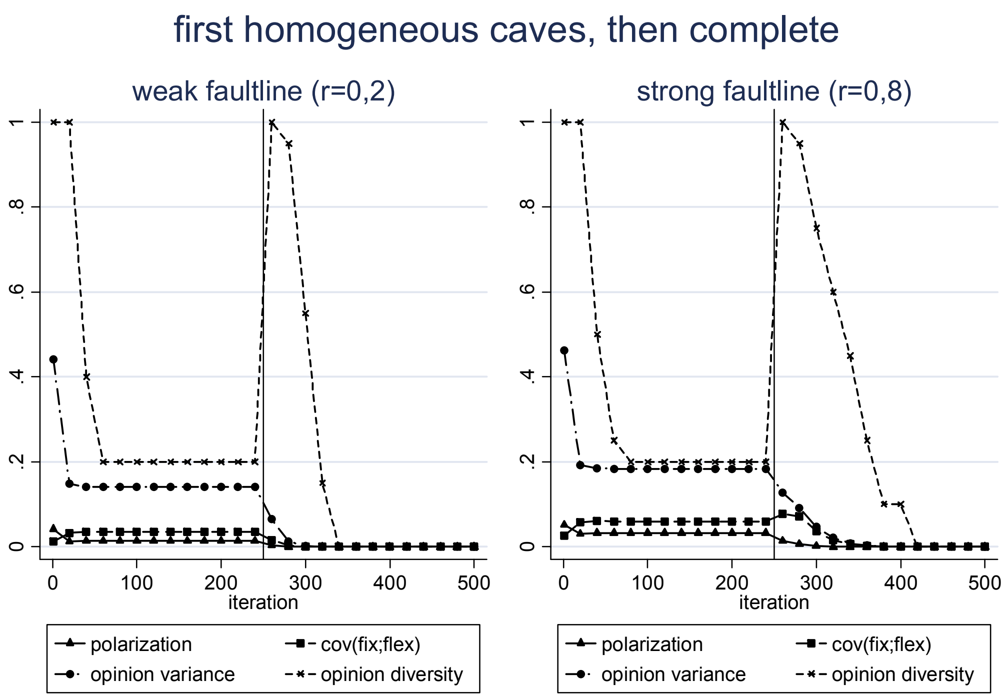
	r = 0.8			r = 0.6			r = 0.4			r = 0.2			r = 0		
i*	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3	attr. A_1	attr. A_2	attr. A_3
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
4	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
5	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1
6	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	1	-1
7	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	1	-1	-1	1	-1
8	-1	-1	-1	-1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1
9	-1	-1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	1	1
10	-1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	1	1
11	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1
12	1	1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
13	1	1	1	1	1	-1	1	-1	1	1	-1	1	1	-1	-1
14	1	1	1	1	1	1	1	1	-1	1	-1	1	1	-1	1
15	1	1	1	1	1	1	1	1	1	1	1	-1	1	-1	1
16	1	1	1	1	1	1	1	1	1	1	1	-1	1	1	-1
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

*Note that actors' numbers i in this table do not correspond to those in table II.1

The result of this procedure is shown in table II.2 which shows the composition of the caves depending on the strength of the faultline. If the correlation between the demographic variables is perfect (a case not included in this table) then there are 4 perfect homogeneous caves: two where all actors hold on all attributes the value +1 and two where all actors hold on all attributes the value -1. The grey cells in table 2 indicate that for other cases some caves are not perfectly homogeneous. While this can not be avoided under the assumptions that $N = 20$ and $N_C = 5$, the table also shows that our method generates a high level of homogeneity within caves.

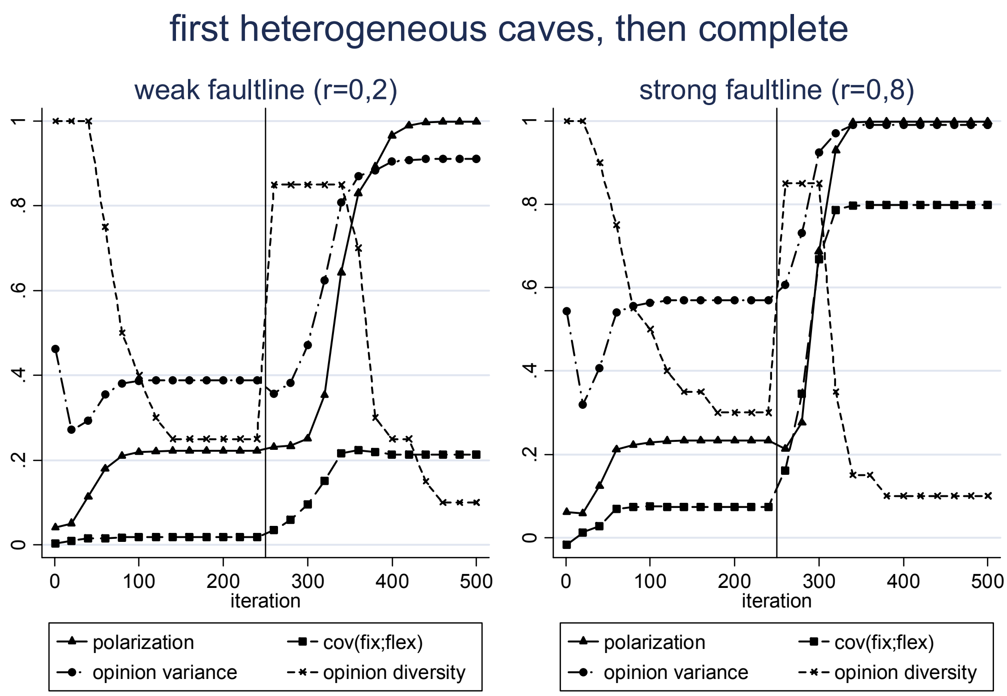
Our method assures that in the condition "first homogeneous caves, then complete", there are almost no negative weights within caves in the initial condition, regardless of the level of faultline strength. Finally, we assumed that in the conditions with caves, the caves are merged in iteration t*=250. This choice of the critical time point assured that the dynamics within caves had practically settled down to equilibrium before the caves were joined.

Figure II.3: Change in outcome measure for typical simulation runs with timing "first homogeneous caves, then complete", for weak faultline (left) and strong faultline (right). N=20, D=3, K=4.



We start again with an illustration of the timing effects by a comparison of typical model dynamics as we obtained them for the two different forms of timing with caves, crossed with low faultline strength (r=0.2) and high faultline strength (r=0.8), respectively. Figures II.3 and II.4 chart for all four settings the dynamics of the four outcome measures until equilibrium. Both figures show for both weak and strong faultlines the type of dynamic that we encountered most frequently in the replications we ran for the corresponding condition (cf. Figure II.6). The corresponding "no timing" baseline is visualized by the time charts in Figure II.1.

Figure II.4: Change in outcome measure for typical simulation runs with timing "first heterogeneous caves, then complete", for weak faultline (left) and strong faultline (right). N=20, D=3, K=4.



The Figures show that the merger of the caves at time t^*=250 dramatically changes group dynamics under both forms of timing. But already in the first phase of the simulated group process there are remarkable differences between homogeneous and heterogeneous caves. The results illustrate that the simulated group dynamics in homogeneous caves exhibit a very strong tendency towards perfect consensus both for weak and for strong faultlines before the merger occurs. *Polarization* drops to nearly zero in this phase and the measures for *opinion variance* and *opinion diversity* approach low levels (about 0.1 and 0.2, respectively). The explanation for this pattern is that homogeneous caves generate local convergence within the caves. Between the agents in a homogeneous cave, there are almost no negative ties. Accordingly their opinions converge towards the average of the randomly chosen initial local opinion distributions. In other words, in each cave the agents reach consensus on moderate opinions. All local initializations are drawn from the same random distribution. As a consequence, the remaining *opinion diversity* and *opinion variance* between the caves is also relatively small. The *opinion diversity* of 0.2 at *t*=250* in both subfigures of Figure II.3 show that exactly four different opinion vectors remain after the early phase, one per cave. The corresponding low *opinion variance* (about 0.1) indicates that the differences between the caves are also very small. By contrast, Figure II.4 shows that in heterogeneous caves groups tend to develop a higher level of *polarization* already within the caves, at both levels of faultline strength. The *polarization* measure increases in both conditions to about 0.2 and *opinion variance* moves to 0.4 (weak faultline) and 0.6 (strong faultline). *Opinion diversity* takes in both runs the value 0.25 before the merger. This indicates that there are 5 different opinion vectors in the team what shows that there is perfect consensus in 3 caves and perfect polarization in one[10]. The local polarization is triggered by the relatively high proportion of negative within-cave ties that is generated due the high likelihood that demographically strongly dissimilar agents are matched within the same cave by the random assignment procedure.

The different developments within the caves set the stage for the dynamics that unfold after merger. A comparison of Figures II.3 and II.4 shows that after *t**=250, groups move to perfect consensus when caves were homogeneous, while perfect polarization is the outcome when initially caves were heterogeneous. Under homogeneous caves, all caves reached consensus on moderate opinions. As a consequence there are virtually no negative ties in the overall group at the point when caves are merged. Accordingly, social influence is overwhelmingly positive and all agents move towards and converge upon the current average group opinion. By contrast, with heterogeneous caves, the dynamics of rejection drove the agents of one cave to the very extreme ends of the opinion dimensions already before the caves are joined. After the merger these extremists exert influence on all team members, with many of whom they have negative ties due to their large opinion differences. As a consequence, agents sufficiently disagree with each other within many dyads, to generate a large proportion of negative ties within the group as a whole at the point when the caves are connected. The result is that in the runs shown by Figure II.4, the previously uncoordinated local division lines merge after *t**=250 into a developing global opinion division, as exhibited by the maximum level of polarization (1.0) shown for the

[10] It is also possible that this result obtains when the opinions in more than one cave polarized and the opinion vectors in the different caves happened to be equal. We checked to make sure that this was not the case in the runs that are reported in Figure V.s III.3 and III.4.

final state in both subgraphs of Figure II.4. The dynamics of the attribute-opinion association *cov(fix;flex)* in Figure II.4 also show that this division occurs mainly along demographical differences when faultlines are strong, whereas the division is only weakly related to demographical differences when faultlines are weak.

For statistical reliability, we conducted again a large number of replications of this simulation experiment and varied faultline strength across the entire interval between r=0 and r=1.0 in steps of 0.2. Figure II.5 reports the results we obtained in both timing conditions for the outcome measures of *polarization* (a), *opinion variance* (b) and *cov(fix;flex)* (c). For comparison, we also include the baseline results of "no timing" in the figures. Results are averages based on 500 replications per condition, where we measured the outcomes after 1000 iterations.

Figure II.5 confirms the patterns exhibited by the typical simulation runs shown in Figures II.3 and II.4. Overall, we find that the indicators for polarization and its association with demographic differences are dramatically lower when the "right" form of timing is chosen (homogeneous caves) than in any of the two alternative cases (no timing or heterogeneous caves). The results also support our intuition that the "right" form of timing strongly tempers the negative effects of faultline strength that we found in the baseline condition of "no timing". As part (a) of Figure II.5 shows, without timing, average *polarization* increases from zero at r=0 (no faultline) to almost the theoretical maximum of 1 at r=1 (maximally strong faultline). With homogeneous caves, there is only a slight increase of *polarization* between those two extremes, from zero at r=0 to 0.2 at r=1. An inspection of the measure of association between fixed and flexible attributes (part (c) Figure II.5) reveals that the formation of homogeneous caves also greatly reduces the degree to which opinion differences in the team align with demographic differences. While the association measure increases for no timing from *cov(fix;flex)* about 0 at r=0 to about 0.6 at r=1, the association measure increases only slightly under homogeneous caves, from *cov(fix;flex)* about 0 at r=0 to about 0.2 at r=1.

Figure II.5: Effect of timing and faultline strength on average *polarization* (a), average *opinion variance* (b) and average association between demographic differences and opinion differences (c), based on 500 replications per conditions, outcomes measured after 1000 iterations per replication N=20, D=3, K=4.

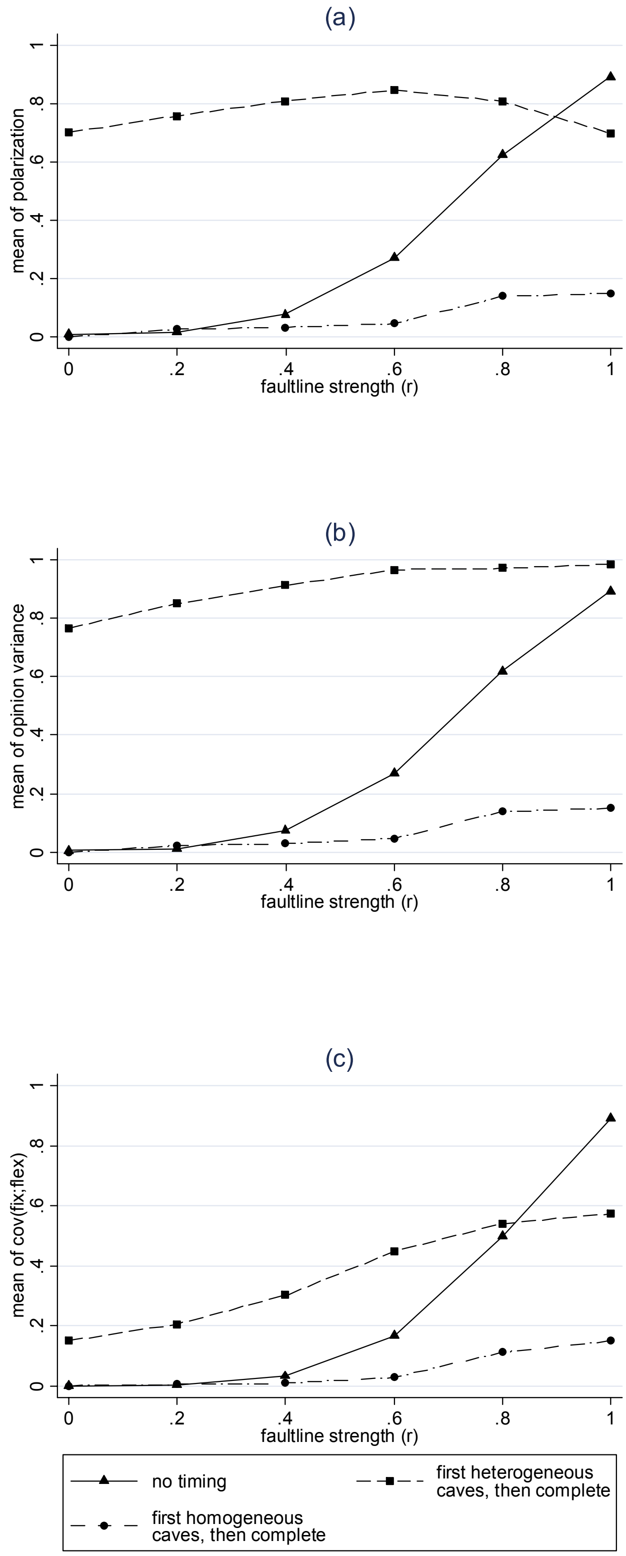


Figure II.5 shows furthermore that the effects of timing strongly depend on the "right" choice for the method of assignment of agents to caves. Broadly, while homogeneous caves generated high levels of consensus and virtually eliminated the negative effects of strong faultlines, heterogeneous caves turned out to make things even worse than they were in the baseline condition of no timing. Part (a) of Figure II.5 shows that with heterogeneous caves, *polarization* was high (about 0.6) even without demographic faultlines and stayed high at all levels of faultline strength. Correspondingly, we found at almost all levels of faultline strength a higher *opinion variance* (part (b)) and stronger association between demographic and opinion differences (part (c)) for heterogeneous caves than for any of the other timing conditions. Only for very strong faultlines (r=0.8 and r=1.0), we find that a further increase in faultline strength is related to a slight decline of the average level of polarization in initially heterogeneous caves, such that for r=1.0 the level of polarization is even somewhat lower than in the baseline condition of no timing. This decline will be explained further below, when we present a detailed analysis of the distribution of equilibrium outcomes that generated the averages reported in Figure II.5.

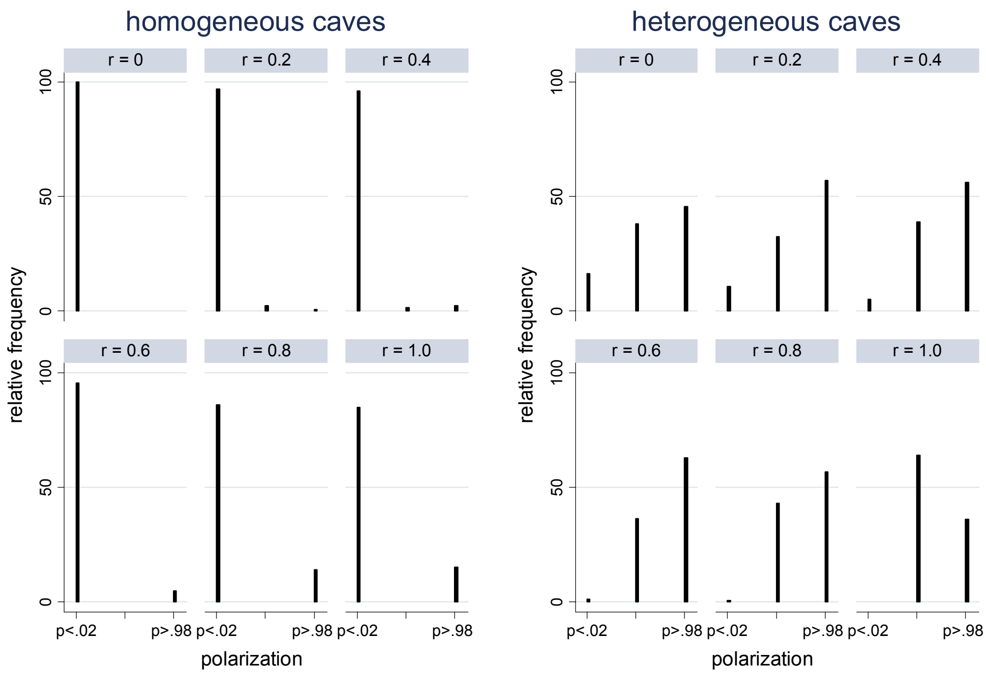
Figure II.5 shows the expected association of the timing conditions with the outcome measures, but it does not directly test our intuition that the effect of the timing conditions can be attributed to a reduction of negative ties in the early phase of the group process. For this, we checked as a first test whether the manipulations of homogeneous and heterogeneous caves affected the proportion of negative ties in the group at the time point before caves were merged (t=250) in the expected direction. While without caves (no timing) on average across all levels of r (3000 runs) 13.3% of all possible dyads were strongly negative ($w_{ij} \leq -0.95$), there was not a single strongly negative tie[11] in any of the 12000 simulated homogeneous caves. As expected, this discrepancy between no timing and homogeneous caves became more pronounced for stronger faultlines, with a maximum level of about 47% of all possible ties in iteration 250 being strongly negative at r= 1.0 with no timing (and zero for homogeneous caves). For comparison, we found 21.9% strongly negative ties at t=250 with heterogeneous caves. These checks confirm our expectation that homogeneous caves suppress the formation of negative ties, while heterogeneous caves foster negativity compared to the baseline of no timing. We also tested whether a higher proportion of negative ties at t=250 was related to higher levels of polarization in equilibrium across all conditions of the experiment. We discovered that even a very small

[11] The number of iterations (250) was chosen large enough to assure that if a tie was negative at this point, it would also be strongly negative, i.e. have a weight w of less than -0.95.

proportion of negative ties at t=250 dramatically increased polarization in the final state. With no timing, average polarization in equilibrium across all levels of faultline strength (3000 runs) was about 0.01 if the proportion of negative ties at t=250 was less than 2.5% (947 runs). The average polarization soared to 0.997 if the proportion of negative ties was above that figure (2053 runs) in this condition. The corresponding figures for heterogeneous caves are an average polarization in equilibrium of about 0.16 if less than 2.5% of the possible ties were negative in t=250 (200 runs), and of about 0.73 with more negative ties (2800 runs). The lower level of polarization for heterogeneous caves reflects the decline of the polarization level that Figure II.5 showed for this condition at high levels of faultline strength r. An explanation for this decline can be found in a closer inspection of the distribution of equilibrium outcomes that generated the averages reported in Figure II.5. We turn now to this analysis.

Figure II.6 shows how timing affects the distribution of *polarization* in equilibrium over the 500 replications that we conducted in each of the two cave conditions in experiment 2. Overall, Figure II.6 supports the interpretation that homogeneous caves greatly increase the odds that a group ends up in consensus, even when faultlines are strong. The highest relative frequency of polarized groups that we obtained with homogeneous caves was 15% for maximally strong faultiness (r=1.0). In almost all other realizations under this timing condition the outcome was perfect consensus. That in some cases groups polarize despite homogeneous caves is a consequence of how the caves were formed. As table II.2 shows, some of the caves are not *perfectly* homogeneous. For r = 0.8, for example, actor 11 holds only on one demographic attribute the same value as the other actors in his cave. The other actors in this cave are demographically completely similar to each other. As a consequence, the likelihood is high that the randomly drawn initial opinions generate differences between actor 11 and his cave mates that are high enough to impose negative ties. Agents will then reject each other's opinions and with a high chance the cave perfectly polarizes. Thus all actors from this cave hold extreme opinions. After the merger those extremists suffice to create enough negative ties in the whole team to make it polarize as well.

Figure II.6: Distribution of polarization measure over 500 replications per condition, broken down by the six different levels of faultline strength and the two different forms of timing. N=20, D=3, K=4.



p < .02: perfect consensus
p > .98: perfect polarization (2 opposing subgroups)
.02 ≤ p ≤ .98: multiplex equilibrium outcome (more than 2 subgroups)

As we expected, the pattern is quite different when caves are heterogeneous. As Figure II.6 shows, we find that even with no faultlines (r=0) about 46% of the replications generated a perfectly polarized group and only about 16% produced perfect consensus. Like for homogeneous caves and for the baseline condition, the stronger the faultline is the less often the dynamics end with perfect consensus, but the overall level of polarization that we find for weak faultlines is much higher. This result differs clearly from the baseline condition (no timing) where about 98% of the runs ended in perfect consensus when there was no faultline. The key factor that drives this result is the relative size of the caves. With heterogeneous caves, the four different caves per group can be seen as four independent replications of the baseline condition, but each with a much smaller group size (N=5 as opposed to N=20) than in the baseline. But the smaller the caves, the more likely it is that there are at least some caves in which there is a relatively high concentration of negative ties from the outset[12]. As a consequence it is relatively likely that the opinions in at least

[12] The reason is that in smaller caves the initial random distribution of opinions produces a relatively sparse coverage of the opinion space. As a consequence, occasional initial "extremists" are likely to have larger

one of the caves polarize perfectly. If the caves are then merged the extremists pull (or push) the rest of the team to the extremes of the opinion scales. Thus the whole team polarizes.

A second main difference between the results from the condition "first heterogeneous caves, then complete" and the other two timing conditions is that with heterogeneous caves multiplex equilibria occur much more frequently. In a multiplex equilibrium, all actors hold extreme opinions, but there are more then only two different opinion vectors in the group. A multiplex equilibrium can arise if the overall pattern of relationships and opinions is exactly balanced so that "push" and "pull" forces exerted upon agents' opinions from different groups of friends and enemies exactly neutralize each other (cf. Macy et al. 2003). Multiplex equilibria are relatively frequent under the timing condition "first heterogeneous caves, then complete" because often the opinions in more than one of the heterogeneous caves polarize before merger. If in addition the opinions of the extremists from different caves differ sufficiently then the extremists pull the moderate actors from the caves that reached a consensus to different poles of the opinion scales after the merger. Table II.3 reports the absolute frequencies of the number of different opinion vectors in teams after 1000 iterations (only for the experiments with heterogeneous caves). If there was only one opinion vector then the whole team reached a perfect consensus. As table II.3 shows, this happened in 81 of the 500 runs with minimal faultline strength. The stronger the faultline the less often the team found a consensus under this timing condition. 2 final opinion vectors indicate that the team perfectly polarized (*polarization* = 1). If there were more than 2 final opinion vectors in the team then a multiplex equilibrium was reached in which all agents hold extreme opinions but there were more then 2 subgroups in the team. This happened when the opinions in more then one of the heterogeneous caves polarized. If there is an unequal number of final opinion vectors larger than 1 then the opinions in more then one cave polarized but the opinion vectors of extremists from two different caves happened to be very similar. As a consequence the members of the two subgroups will have positive relationships. It can thus happen that after the merger these two subgroups of extremists coordinate on the same opinion vector. Table II.3 shows that the higher the faultline strength the more often multiplex equilibria occurred.

opinion distances and thus relatively more negative ties to other group members than in a more densely packed opinion space.

Table II.3: Number of different opinion vectors after 1000 iterations under the condition “first heterogeneous caves, then complete”. (the crosstabulation shows absolute frequencies of runs)

# opinion vectors	Faultline strength						Σ
	0	0.2	0.4	0.6	0.8	1	
1	81	53	25	5	2	0	166
2	286	301	292	316	286	180	1661
3	3	4	14	3	0	0	24
4	99	105	119	128	148	205	804
5	1	2	8	5	7	3	26
6	21	24	28	26	30	53	182
7	3	4	4	2	1	0	14
8	3	6	7	12	26	59	113
9	3	1	2	3	0	0	9
10	0	0	0	0	0	0	0
11	0	0	1	0	0	0	1
Σ	500	500	500	500	500	500	3000

Multiplex outcomes are also the reason why we saw in Figure 5 that the mean of polarization decreases of r=0.8 and r=1.0 and that at r=1.0 the average level of polarization is even lower for initially heterogeneous than for no timing. Figure II.6 makes clear that this result should not be interpreted as showing that there was more perfect consensus under heterogeneous caves in these conditions. For example, in the 500 runs under the condition “first heterogeneous caves, then complete” the dynamics never produced perfect consensus. But Figure II.6 shows that in many runs under this condition more than two groups with partially opposing opinions formed. The opinions of all team members were at the poles of the respective opinion scales in all of these cases. However, the outcome was multiplex, so that there was not perfect polarization into *two* opposed subgroups in these cases. This is reflected by a value of the polarization measure that is somewhat lower than the theoretical maximum of 1.0, but still considerably above the level for consensus (0.0), which explains why on average across all runs we found a polarization level of about 0.7 in this condition.

We believe the reason that we find more multiplex outcomes with heterogeneous caves than in any of the other timing conditions lies with the uncoordinated local polarizations that under strong faultlines are likely to arise at the end of the first phase. In a four dimensional opinion space, there are only 16 possible combinations of extreme positions on every dimension. Equilibrium outcomes will arise if agents are distributed in the right way over these 16 combinations, or over smaller subsets of the combinations (e.g. with 4 or 8 different opinion vectors in the group) such that all mutual influences on agents’ opinions are in balance. With heterogeneous caves, every locally polarized cave produces at least two of those combinations with some incumbents. With five caves overall, it is not unlikely that this process generates an overall distribution in the group as a whole that is in or close to a multiplex equilibrium when all caves are merged.

II.4. Summary and Discussion

We modeled in this paper the effects of demographic faultlines on team performance. Lau and Murnighan's theory suggests that the stronger a team's demographic faultline is the less cohesive the team will be and the less likely will the team therefore be able to find a consensus with regard to work related opinions. As a consequence, teams with a strong demographic faultline tend to perform poorly. We proposed a formal computational model of this process based on four fundamental sociological mechanisms, homophily, heterophobia, social influence and rejection. We showed that the model generates results that are consistent with Lau and Murnighan's faultline theory. Our simulations demonstrate that the stronger the demographic faultline in a group the more likely will the group split up into subgroups (ceteris paribus). These subgroups' members hold opposing opinions and do not like each other.

We then used our model to show that the degree to which strong faultlines have negative effects may critically depend on the timing of contacts between group members. We tested a somewhat counter intuitive prediction: if in the first phase of the team interaction the team is separated into demographically homogeneous groups which are merged only later in the team process, then strong demographic faultlines do less often lead to opinion polarization than in a process where all group members interact with each from the outset. This result contradicts to some extent predictions of the prominent contact theory (Allport 1954; Pettigrew 1998) which states that contact improves the interpersonal relationships between demographically dissimilar actors. However, as discussed above, the effects of timing follow logically from the fundamental social mechanisms that constitute our model.

It was our main interest in this paper to show the theoretical consistence of the reasoning that implies effects of the timing of contacts in demographically diverse groups. Accordingly, we did not conduct an extensive analysis of the robustness of our results with regard to variation in other model parameters than those we have manipulated. We have shown elsewhere (Flache and Mäs 2008b) that our model reconstructs basic predictions of faultline theory also for different numbers of opinions and demographic dimensions. It is a task for future research to conduct more extensive sensitivity analyses. At this point we see no a priori reason to expect that qualitative model results may fundamentally change for other sets of parameters, as long as the parameters of the model are chosen such that the model equations are consistent with the social mechanisms we assume.

The mechanisms we use in our model imply that timing is not the only manipulation that may avoid the negative effects of strong demographic faultlines on team cohesion. More generally, according to our model every condition that suppresses the emergence of negative ties helps to sustain group cohesion despite demographic divisions. Team building measures, emphasis on common goals or team learning may have similar effects than the right form of timing. Such measures have been proposed by previous research on the faultlines (e.g. Gibson and Vermeulen, 2003). Also previous work on agenda setting points to measures that may have similar effects (see e.g.: Levine and Plott 1977; List 2004; Plott and Levine 1978). Team managers might manipulate the sequence in which certain issues are discussed. If in a first phase only salient issues are discussed that all team members agree on, this would imply the emergence of positive interpersonal relationships between the team members. If then more controversial issues are addressed in a later phase, the prospects for finding a consensus are much better than compared to a situation where only controversial topics were addressed from the outset. Clearly, previous research on team building and agenda setting points to fruitful new applications of our model, but we also wish to emphasize that with the manipulation of the timing of contacts that we addressed in this paper our model suggests a measure that to our knowledge is new in the literature. One possible advantage of timing may be that it is a measure that organizations can implement unobtrusively, seemingly as a byproduct of functional arrangements of the workflow.

Future research should also focus on the mechanisms that produce opinion polarization. As we argued above, Lau and Murnighan's reasoning seems to critically hinge upon the assumption that there is an initial correlation between demographic attributes and opinions. In our model, this assumption is not necessary. Instead, the two negative mechanisms of heterophobia and rejection are sufficient to generate an effect of faultline strength on opinion polarization. We propose that future work should compare our model with the Lau and Murnighan reasoning on a theoretical level, to search systematically for contradicting predictions that can subsequently be submitted to empirical tests. We suggest that effects of the timing of contacts are particularly promising to compare the models empirically. We expect that the mechanisms we used and those of Lau and Murnighan produce different dynamics under certain timing conditions. As we have shown, our mechanisms produce less polarization if first homogeneous subgroups are formed. By contrast, Lau and Murnighan's mechanism should lead to the opposite outcome. Their reasoning implies that in homogeneous groups the actors agree on opinions and that their opinions should become more extreme then. Because Lau and Murnighan assume that demographically dissimilar actors also hold opposing opinions, each of the subgroups will find a different very extreme consensus. If then the team is merged, all team members hold very extreme opinions and a consensus is very unlikely. If on the other hand in the first phase heterogeneous groups are formed then the actors in each subgroup will hold different opinions. If they then exchange the arguments their opinions are based on they may be able to convince each other. From this view, it is thus very likely that the subgroups find a consensus on moderate opinions. After the merger, the moderates will very likely find an overall consensus. Hence, the predictions of our model and the Lau and Murnighan reasoning are contradictory under certain timing conditions.

Our analysis has demonstrated how the theory of faultlines can be rigorously and formally reconstructed. We also have shown that this reconstruction can yield new, empirically testable hypotheses into the conditions and mechanisms that may temper or elicit the negative effects of demographic faultlines on team performance. Finally, our analysis suggests that the timing of contacts is a potentially fruitful governance instrument that managers may be able to use in order to avoid that the negative effects of demographic faultlines overshadow the benefits that diverse human and social capital can create for organization.

III. The Polarizing Effects of Argument Exchange[13]

Abstract
In the previous chapter, we analyzed a social-influence model that includes the assumption of negative influence and demonstrated that this model can generate opinion polarization. However, in the following we will show that empirical research provided mixed evidence for negative influence. For this reason, we will develop in this chapter an alternative theory of opinion polarization.

Our new theory explains polarization (denoted bi-polarization here) without the negative-influence assumption. This approach assumes that opinions are based on arguments and that these arguments are exchanged during interaction. When individuals with similar opinions interact, they likely provide each other with new arguments that support their opinions. In this way, opinions become more extreme. We will propose that in combination with homophilious selection of interaction partners, the exchange of arguments can lead to polarization of opinions.

We will proceed in two steps. First, we will develop a formal model of opinion dynamics that is based on argument exchange and demonstrate that this model can indeed generate polarization. A computational experiment will reveal that *strong* homophily is a crucial condition of polarization in this model.

Second, we have put the new theory to the test. In a laboratory experiment, we tested our central proposition that argument exchange can cause polarization. Groups of 8 participants discussed an issue and could either exchange arguments or only opinions. We found polarization, but as predicted only when arguments were transmitted and when there was strong homophily in the selection of interaction partners.

III.1. Introduction

Sociological and socio-psychological theories of intergroup processes, like social differentiation (Bourdieu 1984[1979]; Mark 2003), ingroup favoritism (Tajfel 1981), outgroup discrimination (Mummendey et al. 1999), and intergroup conflict (Sherif 1966; Tajfel and Turner 1986) build on the assumption that individuals seek to accentuate differences between their own group and salient outgroups. Sociological approaches, for

---

13 This chapter is co-authored with Andreas Flache and is submitted for publication in a social psychological journal. Note that we adjusted the terminology. In particular, we refer to the development of clusters with increasingly distant opinions as 'bi-polarization' and not 'polarization'. This was necessary because the term 'polarization' describes collective extremization tendencies in the social psychological literature.

instance, argue that people develop elaborated cultural norms in order to distinguish themselves from groups with a lower status (Bourdieu 1984[1979]; Bryson 1996; Elias 1969[1939]; Simmel 1957; Turner 1995). In the same line of reasoning, psychological theories building on the self-categorization paradigm (Brewer 1991; Tajfel and Turner 1986; Turner 1987) hold that humans adjust their opinions and behavior in a way to minimize the heterogeneity within their ingroups and to maximize differences to outgroups (Hogg, Turner and Davidson 1990).

The notion that humans seek to accentuate intergroup differences has recently been incorporated into theories of opinion dynamics (Baldassarri and Bearman 2007; Salzarulo 2006). In particular, it has been assumed that in an intergroup context, individuals may adapt their opinions in order to maximize disagreement with a perceived outgroup opinion. With two groups perceiving each other as outgroup, this suggests *bi-polarization*, the development of a bimodal opinion distribution in the course of social interaction, with gradually increasing distance between the opposite modes. In fact, empirical studies of opinion dynamics have provided some evidence of bi-polarization tendencies with regard to salient opinions, for example among college students (Feldman and Newcomb 1969) or in ethnically mixed work teams (Early and Mosakowski 2000). In a similar vein, observers of the dynamics of political opinions found tendencies towards bi-polarization on controversial issues in the American public during election periods (Abramowitz and Saunders 2008; Evans 2003; Fiorina and Abrams 2008; Levendusky 2009). Existing sociological and socio-psychological theories of bi-polarization explicitly assume what we denote here "negative influence", the tendency of individuals to adjust their opinions in a way to increase opinion differences to dissimilar others (Baldassarri and Bearman 2007; Hogg, Turner and Davidson 1990; Macy et al. 2003; Salzarulo 2006). Yet, as will be elaborated in more detail further below, empirical research on opinion dynamics provided mixed evidence for negative influence and has been criticized on methodological grounds (e.g. Krizan and Baron 2007).

This raises the question how bi-polarization might be explained without the assumption that individuals seek to increase opinion differences to dissimilar others. In this paper, we present and test a theory of bi-polarization that does not rely on negative influence. Moreover, our theory does also not need to assume that bi-polarization results from the interaction between two distinct groups who regard each other as salient outgroup. Instead, we model bi-polarization as consequence of an *intra*-group process.

Our theory has two main ingredients. First, we draw on Persuasive Argument Theory (PAT) (Myers 1982; Vinokur and Burnstein 1978). PAT holds that individuals base their opinions on arguments and influence each others' opinions when they exchange arguments. When individuals with similar opinions exchange arguments, they may provide each other with new arguments that support their opinions. As a consequence, their opinions may be intensified and become more extreme. We combine PAT with the assumption of *homophily* (Lazarsfeld and Merton 1954) (Ibarra 1992; Lazarsfeld and Merton 1954; McPherson, Smith-Lovin and Cook 2001; Moody 2001), or more specifically the notion that individuals tend to interact with others who hold similar opinions. Sociological research has shown that homophily, or the tendency of "birds of a feather to flock together" is a robust empirical regularity with regard to similarity in a wide range of characteristics, including opinion similarity (Byrne 1971; McPherson, Smith-Lovin and Cook 2001).

In the theory section of this paper we present an informal reasoning that describes how the interplay of persuasive argument exchange and homophily may give rise to bi-polarization. However, without precise modeling it is hard to gain solid intuitions about the social outcomes ensuing from simultaneous interactions of multiple individual group members driven by these two principles. Recently, an increasing number of theorists advocate employing agent-based computational modeling (Bonabeau 2002; Macy and Willer 2002; Smith and Conrey 2007) to understand the "large-scale consequences of the theoretical assumptions about individual behavior when the behaviors are carried out in the context of many other agents and iterated dynamically over an extended period of time" (Smith and Conrey 2007: 88). To assure the logical consistency of our reasoning, we therefore elaborate a computational agent-based model that shows how and under what conditions the interplay of persuasive argument exchange and homophily can generate bi-polarization.

Furthermore, we put the theory to a test. We conducted a laboratory experiment (N=96) with a controlled group discussion process in groups of 8 participants. The focus was on testing how bi-polarization is affected by the two key mechanisms, argument exchange and homophily. For this, we manipulated independently the possibility for participants to exchange the arguments on which they build their opinions (rather than, or in addition to being only exposed to each others' opinions) and homophily, operationalized as the extent to which interaction was restricted to pairs of participants with similar opinions. We carefully avoided that participants received information on which to base social categorizations of others into in- our outgroup members. This precluded bi-polarization driven by the mechanism of negative influence. We hypothesized that bi-polarization would nevertheless occur, but only in the experimental condition in which participants could exchange the arguments underlying their opinions, and in which homophily was imposed in the matching of interaction partners.

III.2. The critical role of negative influence

Recent contributions to the literature on social-influence dynamics (for recent review see Mason, Conrey and Smith 2007) demonstrate the critical role of the negative-influence assumption in theories of bi-polarization. Early formal models failed to generate bi-polarization (Abelson 1964; Berger 1981; French 1956; Harary 1959; Wagner 1982) because only positive influence was assumed. Instead, these models imply that a group inevitably ends up in consensus as long as there are no subgroups that are entirely cut off from outside influences. In search for processes that give rise to bi-polarization, an increasing number of models have therefore been proposed that combine both positive influence from similar and negative influence from dissimilar sources (Baldassarri and Bearman 2007; Flache and Mäs 2008a; Macy et al. 2003; Mark 2003; Salzarulo 2006).

These theoretical accounts of bi-polarization hinge critically upon the assumption of negative influence. However, experimental tests have hitherto not provided unequivocal evidence in support of this assumption. In laboratory experiments, researchers typically informed participants about the opinions of fictitious members of both ingroup and outgroup and then measured pre-test–post-test opinion shifts. These studies have led to very mixed results. Many did not find increasing differences between in- and outgroup opinions at all (Hogg, Turner and Davidson 1990; Krizan and Baron 2007; Lemaine 1975). In addition, research illustrates that individuals may publicly distance themselves from others but their private opinions actually do not shift (Berger and Heath 2008).

Moreover, methodological issues cast doubt on the conclusiveness of those studies that researchers interpreted as support for negative influence (Berscheid 1966; Hogg, Turner and Davidson 1990; Mazen and Leventhal 1972; Sampson and Insko 1964; Schwartz and Ames 1977; van Knippenberg and Wilke 1988; van Knippenberg, De Vries and van Knippenberg 1990). Krizan and Baron (2007) raised a number of issues with regard to experiments in the group polarization tradition. We point here to two major additional problems. First, some experimental designs do not allow to disentangle positive influence from the ingroup and negative influence from the outgroup in the explanation of opinion shifts (e.g. Hogg, Turner and Davidson 1990; van Knippenberg and Wilke 1988; van Knippenberg, De Vries and van Knippenberg 1990). In these studies, participants have been exposed two sources of social influence, ingroup members and outgroup members. Participants were exposed to ingroup members who held opinions relatively similar to their own. Some of these ingroup members held more extreme opinions than the participant. Outgroup members always had opinions distinct from those of the participants. With such a design, opinion changes away from the outgroup opinion may have been caused by both negative influence from the outgroup or positive influence from more extreme ingroup members (Mackie 1986).

The second problem is that some studies did not control for general opinion drifts during the experiment (Mazen and Leventhal 1972; Sampson and Insko 1964). For example, Mazen and Leventhal (1972) confronted expectant mothers with a favorable description of breast feeding and measured how this affected the mothers' opinions on this issue. They found that mothers developed more positive opinions when they received information from a communicator with a similar skin color (positive influence). But, when the communicator and the mother were dissimilar in skin color the opinions of the mothers turned more negative. This suggests that these mothers were influenced negatively by the communicator. However, we argue that this result may have been caused by a general trend towards more negative opinions. In this study, the second opinion measurement took place one week after the first. In this period, participants might have developed more negative opinions. However, those mothers who where similar to the communicator were positively influenced by them and changed their minds back to more positive opinions. The opinions of the dissimilar mothers, however, might have been unaffected by the communicator's information and remained more negative. Unfortunately, the authors did not control for trend effects in their analyses. It is therefore not clear whether the reported opinion dynamics are the result of negative influence or opinion drifts.

In sum, existing theories of bi-polarization critically hinge on the assumption that individuals tend to adjust their opinions in a way to increase opinion differences to dissimilar others. However, there is hitherto no conclusive empirical evidence for negative influence. In the following section, we elaborate a theory of bi-polarization which does not rely on negative influence.

III.3. Theory

Our theory of bi-polarization builds on earlier theorizing on demographic faultlines (Lau and Murnighan 1998) and group polarization (Myers 1982; Myers and Bishop 1970). These approaches already combined insights from PAT (Isenberg 1986; Vinokur and Burnstein 1978) and research on homophily (Ibarra 1992; Lazarsfeld and Merton 1954; McPherson, Smith-Lovin and Cook 2001; Moody 2001).

PAT has been developed to explain why discussion groups which initially have a tendency towards one side of an issue will become more extreme in their opinions as a result of discussion. This phenomenon, called *polarization*, has been robustly demonstrated by a range of experimental studies (Isenberg 1986; Myers 1982). PAT assumes that individuals base their opinions on pro and con arguments. During discussion, individuals are exposed to the arguments their interaction partners consider relevant. In groups where members tend towards a specific opinion already prior to discussion, mainly those arguments will be brought up that favor the prevailing tendency. Discussion members, thus, provide each other with further arguments that support their initial position. This intensifies opinions and aggregates to a *collective* opinion shift towards more extreme positions.

Building on earlier work (Lau and Murnighan 1998; Myers 1982; Myers and Bishop 1970) we argue that the interplay of the persuasive argument exchange described by PAT with homophily can give rise to bi-polarization. The idea is that small initial opinion differences in a group are gradually amplified when argument exchange occurs more frequent between those individuals who initially have relatively similar opinions than between those whose opinions are relatively dissimilar. Due to homophily, individuals with opinions leaning towards the same pole of the opinion spectrum interact more likely with each other than with those who lean towards the opposite pole. Thus, persuasive argument exchange reinforces existing opinion tendencies, but in opposing directions in the separate subsets of group members who share the same initial tendency. This further reduces the likelihood of interaction between initially dissimilar pairs of individuals, which in turn further strengthens existing tendencies. This process unfolds simultaneously at both sides of the opinion spectrum, such that a self-reinforcing dynamic may arise that entails bi-polarization even in the absence of negative influence.

Bi-polarization requires homophily according to this reasoning. However, it remains unclear how strong homophily needs to be to render bi-polarization a likely outcome of the dynamic. As long as there is some probability of interaction also between actors with dissimilar opinions, bi-polarization tendencies might be very unlikely. When actors with dissimilar opinions interact, they likely exchange arguments that speak against their current tendency and lead to more moderate opinions. Furthermore, in subsequent interaction actors will transmit these counter arguments to similar others. This will lead to further opinion convergence. In sum, this reasoning suggests that even though actors may tend to interact with similar others, occasional deviations from this rule may suffice to impede the bi-polarization tendencies of argument exchanged and homophily. In a non-deterministic world, there is no guarantee that a self-reinforcing dynamic eventually leads a social system into the state towards which the dynamic moves. This has for example been demonstrated for formal stochastic models of residential segregation (e.g. Stauffer and Solomon 2007), or cultural dissemination (e.g. Klemm et al. 2003a). In these models, the "ordered" outcomes towards which individual decision rules drive the system, such as highly segregated residential distributions, or local clustering of similar cultures, only arise when the level of randomness in individual decision making is relatively small.

To identify how strong homophily needs to be to give rise to bi-polarization, we developed and applied an agent-based computational model of our theory. In the following section, we present the model. Subsequently, we report results from a computer simulation experiment designed to assess the relationship between the strength of homophily and bi-polarization.

III.3.1. The formal model

The agent-based model implements the substantive assumptions of PAT and homophily for each of N interdependent individuals who simultaneously participate in an artificial influence process. Each individual is represented as an agent i, with a numerically valued opinion o_i $(-1 \leq o_i \leq +1)$ that represents the agent's stance on a given issue. We assume that there is a limited number of arguments that address the issue. The valence of an argument is expressed numerically. More precisely, P pro arguments ($a_l = 1$) and C con arguments ($a_l =$ -1) are available. This is summarized in the argument vector, an array of arguments with P+C elements. Elements with a row number smaller than P+1 hold pro arguments, i.e. $a_l = +1$. The remaining elements contain con arguments, i.e. $a_l = -1$.

Empirical research suggests that people have limited capacities to remember and process information (Cowan 2001; Miller 1956). Accordingly, we assume that agents base their opinion only on a subset of S relevant arguments ($S \leq P+C$). The remaining arguments are not relevant in the opinion formation. Technically, an agent's opinion is the average value of the arguments a_l that the agent considers relevant (see equation 1). For simplicity, we assume that all relevant arguments have the same persuasiveness. Technically, this is expressed by the assumption that all relevant arguments are equally weighted in the calculation of the opinion.

$$o_i = \frac{1}{S}\sum_{l=1}^{S} a_l \qquad (1)$$

For example, an agent i that bases her opinion on 6 pro arguments (S=6) holds an opinion of o_i=1. However, if the agent considers e.g. 3 pro and 3 con arguments relevant, the opinion will take the value zero.

Following research on memory processes (Brown and Chater 2001), we assume that agent's disregard pieces of information if they are not sufficiently recent. Thus, the more recent an argument is at a given point in time, the longer this argument remains relevant for the formation of the agent's opinion. This is implemented for each agent in a recency vector. This vector has P+C elements. Each element indicates how recent the respective argument is for the agent. Elements of the recency vector with a row number smaller than P+1 identify the relevance of pro arguments. The remaining elements determine the relevance of con arguments. Arguments are either relevant or not, but agents rank the S relevant arguments according to their recency. We denote the recency of an argument ($s_{l,i}$) with integer values between 0 and S ($s_{l,i} \in \{0,...S\}$). A value of $s_{l,i}$ = 0 indicates that the argument a_l is *not* sufficiently recent and therefore *not* relevant for actor i. Values above zero indicate that this argument *is* sufficiently recent and therefore affects actor i's opinion. The most recent argument has the value of $s_{l,i}$ = S, the second most recent argument has the value S-1, and so on. Thus, if an agent considers three arguments (S=3) then one has a recency of 1, one has a recency of 2, and one has a recency of 3. The recency rank of an argument does *not* affect the extent to which an argument shapes the current opinion (see equation 1). However, the recency determines *how long* an argument affects the agent's opinion in the influence process. The exact rules for updating argument-recency will be elaborated further below.

We model the opinion formation process as a sequence of events, each event corresponding to one interaction between two agents. An interaction consists of a partner selection phase and a subsequent social influence phase. In the partner selection phase, two agents from the population are matched for interaction, based on opinion-homophily.

Subsequently, an opinion of one of the interacting agents is updated as a result of the interaction. The updating rule operationalizes the argument exchange mechanism of PAT.

We implement the *partner selection* phase as follows. In each event, the computer first randomly picks an agent i^*. Then an interaction partner j ($j \neq i^*$) is selected. The probability that agent j is chosen as interaction partner depends on the similarity between i^* and j, $sim_{i^*,j}$, that varies between 0 and 1. A similarity of zero expresses maximal dissimilarity, whereas $sim_{i^*,j} = 1$ if both actors hold exactly the same opinion. Formally,

$$sim_{i^*j} = \frac{1}{2}\left(2 - \left|o_{i^*} - o_j\right|\right) \tag{2}$$

The probability that agent i^* chooses j as interaction partner (p_j) derives from their relative similarity, that is: the degree to which j is more similar to i than other group members are. Technically,

$$p_j = \frac{\left(sim_{i^*j}\right)^h}{\sum_{j=1, j \neq i^*}^{N}\left(sim_{i^*j}\right)^h} \tag{3}$$

Equation 3 implements homophily. The more similar j is to i^* the higher is the probability that they will interact. If two actors differ maximally then the probability of interaction equals zero. To vary the *strength of homophily* we include the parameter h into the model. The higher the value of h, the steeper is the increase of the likelihood that j will be chosen by i^* as an interaction partner in the relative similarity of i^* and j. The actual selection of the interaction partner of i^* is implemented by a random draw of one agent from the set of all other group members, based on the probabilities p_j given by (3).

Next, i^* is socially influenced by the selected interaction partner j^* based on the persuasive arguments mechanism. For this, the computer randomly picks one argument, a_{l^*}, out of the S arguments that j^* considers relevant. Each relevant argument has the same probability to be chosen ($1/S$). Arguments that are not relevant for j^* are not chosen. The chosen argument is then adopted by i^* . Technically, its recency for i^* is updated to a value of S+1 ($s_{l^*,i^*} = S+1$). Subsequently, the recency of all arguments that have non-zero recency in i^*'s recency vector is reduced by one, if prior to the interaction the corresponding argument was more recent for i^* than the argument adopted from j^*. As a result, the argument that was communicated by j^* becomes relevant for i^* and attains the highest recency of all argument that i^* considers relevant ($s_{l^*,i^*} = S$).

This updating procedure implements the assumptions about agents' limited capacity to memorize information (Cowan 2001; Miller 1956) and their bias towards considering only recent information (Brown and Chater 2001). It implies in particular that agents "forget" one of the arguments previously relevant for them, if they have learned a new argument in the interaction. This assures that the number of arguments that is relevant for an agent is kept constant at S throughout the influence process.

Interaction events are iterated until the system reaches equilibrium. Our model has exactly two equilibria, perfect consensus and maximal bi-polarization. Perfect consensus is reached when all agents hold the same opinion and base it on the same set of arguments. Perfect consensus is a stable situation because agents can transmit only arguments which their interaction partners already consider relevant. This implies that opinions will not be affected by argument exchange. Maximal bi-polarization obtains if there are two maximally distinct subgroups and the members of each subgroup agree on opinions and arguments with each other. That is, the members of the subgroups have coordinated on the opposite poles of the opinion scale and the pairwise similarity ($sim_{i,j}$) between agents of different subgroups is zero. In this situation, the probability is zero that agents interact who belong to different subgroups (see equation 3). Argument exchange between the subgroups is thus precluded. In addition, interaction of agents that belong to the same subgroup can not lead to opinion changes because these agents base their opinion on either exclusively pro-arguments or exclusively con- arguments. Any outcome of the process that is not perfect consensus or perfect bi-polarization can not be an equilibrium. The reason is that any other outcome implies that there are differences in opinions or arguments between agents, and a positive probability of interaction between the agents who hold different opinion or arguments. There is thus a positive probability that the distribution of arguments and opinions in the population will change due to interaction.

III.3.2. Dynamics of Bi-polarization: an illustrative simulation run

We began by testing whether the model can generate bi-polarization. For this, we imposed conditions for which we expected bi-polarization tendencies to be very strong. Accordingly, we imposed relatively strong homophily, assuming h=9. With this value, homophily is so strong that interaction between agents who do not hold perfectly similar opinions is extremely unlikely. Furthermore, we assumed that thirty pro and con arguments are available (P=C=30) and all agents consider 10 relevant arguments at the same time

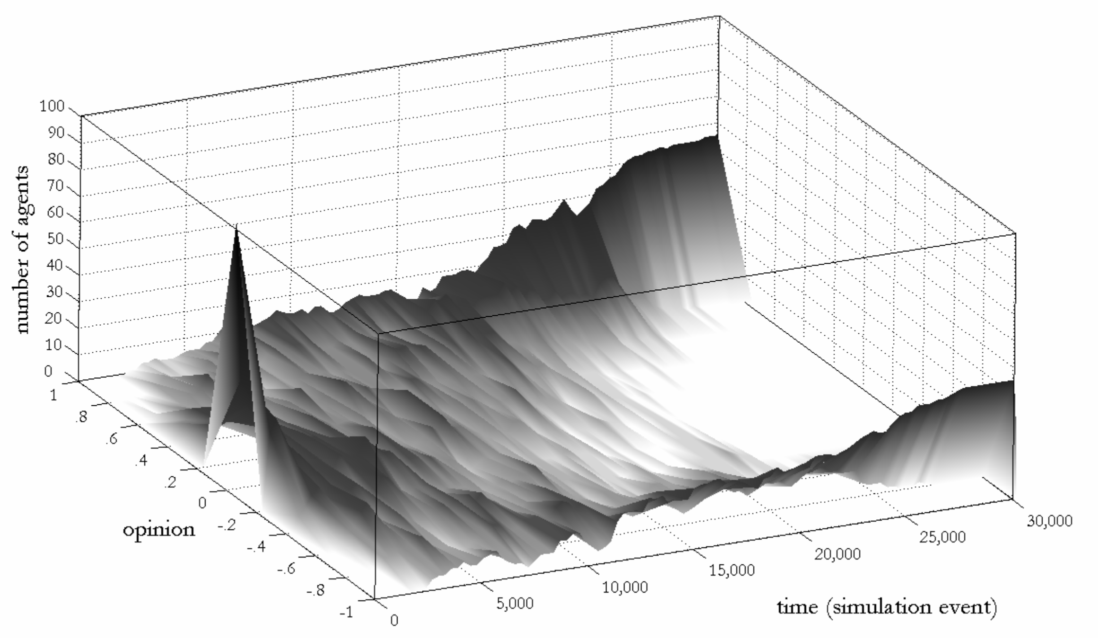
(S=10). For this condition, we simulated a population of 100 agents and studied the change of agents' opinions and argument vectors over 30,000 simulation events.

The initial distribution of arguments and opinions was created by assigning to each agent a random set of 5 pro and 5 con arguments. With this, all agents started with the same opinion at the middle of the opinion scale (o_i=0 for all i). Thus, at the outset there are no opinion differences between agents. Notice that this does not rule out opinion changes, as there are still differences between agents in the arguments on which their opinion is based.

Figure III.1 shows a surface which depicts the development of the opinion distribution during the typical simulation run. The shading of the surface's areas and the respective value on the z-axis indicate how many agents hold a certain opinion at a certain simulation event. White areas indicate that no agent holds the respective opinion. The darker the area, the more agents hold this opinion. At the beginning of the simulation (event zero), all 100 agents hold the same opinion. The figure shows how bi-polarization emerges in this simulation run. While opinions are approximately uniformly distributed after about 10,000 simulation events, the distribution becomes bimodal after about 15,000 events. Subsequently, the two modes gradually become more accentuated and shift towards the opposite ends of the opinion spectrum until, after about 30,000 events, the population is almost entirely split into two approximately equally large subsets of agents with opinions of -1 and +1, respectively.

Opinion change is possible despite initial uniformity, because agents base their opinion on different (randomly assigned) sets of five pro and five con arguments. Thus, in some interactions agents' opinion shifts away from the initial consensus, because they learn a new pro (con) argument and forget a con (pro) argument. Their new opinion is then based on more pro (con) than con (pro) arguments and takes a positive (negative) value. Figure III.1 shows that this results in an increase of the variance of the opinion distribution in the first phase of the simulation run. After about 10,000 simulation events, the opinion is uniformly distributed. Due to the strong homophily, agents are matched with interaction partners that have adjusted their opinion in the same direction. These interaction partners will more likely provide each other with arguments that further intensify their opinion tendency rather than to communicate arguments that render their opinions more moderate again. Eventually the opinion trajectories of all agents move to one of the two outer ends of the opinion scale. At this point, the opinion distribution stabilizes, because agents base their opinions on either only pro or only con arguments such that interaction is only possible between agents who already hold identical opinions. Agents can no longer learn arguments that could change their opinions.

Figure III.1: Bi-polarization generated by argument exchange and homophily (N=100, P=C=30, S=10, h=9)



To summarize, this illustrative simulation run confirms that the interplay of argument exchange and homophily can generate bi-polarization even in the absence of negative influence. What is more, in this run bi-polarization emerged even though we assumed perfect opinion consensus at the outset. In sharp contrast, existing models of continuous influence dynamics (Baldassarri and Bearman 2007; Flache and Mäs 2008a; Macy et al. 2003; Salzarulo 2006) imply that bi-polarization can only arise when there are initially differences between members of a population which can form the basis of group categorization and negative influence.

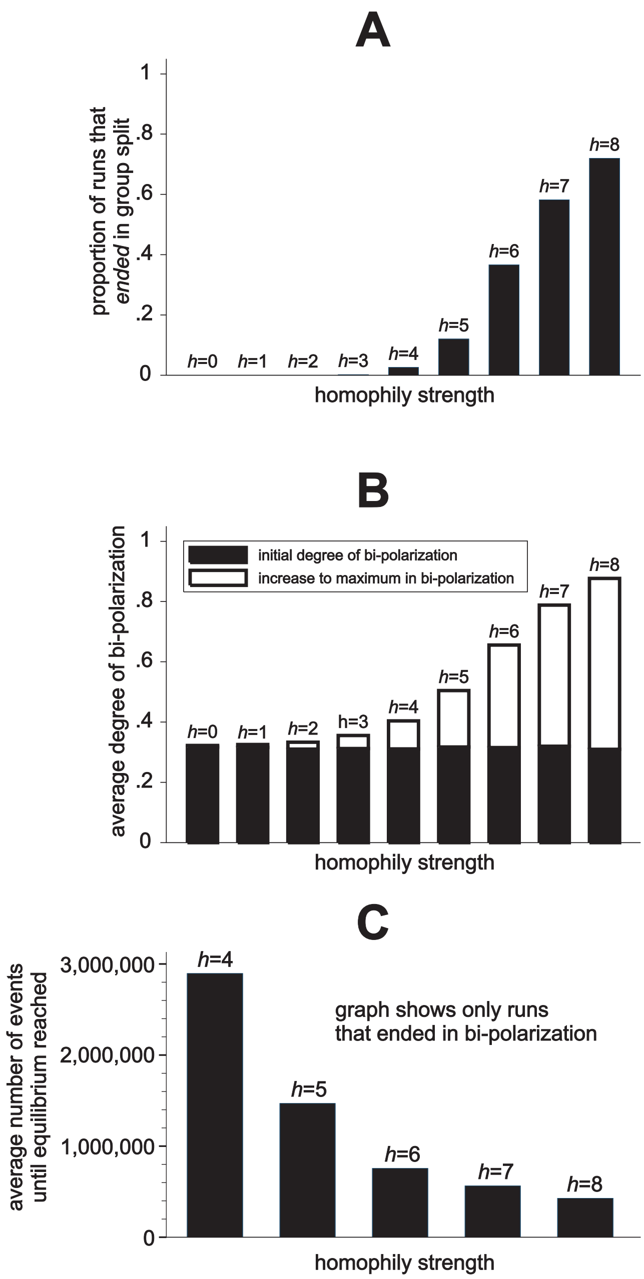
III.3.3. Effects of Homophily

Next, we wanted to know whether homophily always entails bi-polarization, or whether bi-polarization can only arise when homophily is sufficiently strong. We conducted a simulation experiment in which we varied the model parameter h between 0 (no homophily) and 8 (strong homophily) in steps of 1. Per condition, we ran 500 independent replications of the simulation. In all simulations of this experiment, we studied populations of 20 agents (N=20). This is a plausible group size for school classes and work teams (see e.g. Wegge et al. 2008), two of the settings for which theory and empirical accounts of intra-group conflicts suggest the possibility of bi-polarization dynamics (see e.g. Lau and Murnighan 1998). We assume that there are 20 pro and 20 con arguments (P=C=20) available, and that agents can only take 6 arguments into account simultaneously (S=6). Values for P, C and S were selected to create sufficient variation in the initial argument sets also between agents who hold identical opinions. For this, P and C should considerably exceed S. Otherwise agents with similar opinions very likely base their opinions on similar sets of arguments. This would preclude the possibility that argument exchange between agents with similar opinions renders their opinions more extreme because they provide each other with arguments which they already consider relevant. Furthermore, we created the initial condition such that opinions are uniformly distributed. For this, we randomly assigned to each agent one of the S+1 possible opinion values and then randomly picked one of the possible sets of S arguments which correspond to the selected opinion value.

Figure III.2 summarizes the results. Panel A of figure III.2 shows how homophily strength h affected the proportion of runs that ended in bi-polarization. When homophily strength was below h=3 all runs ended in consensus. At h=3, only one out of the 500 replications for this condition ended in a group split with two subgroups at the opposing poles. For higher values of homophily strength h, panel A shows that the stronger homophily was, the more runs ended in a perfect group split.

If a simulation runs ends in perfect consensus, there may nonetheless have been a temporary period of significant bi-polarization in the dynamic (see chapter IV). To test for this possibility, we assessed for each simulation run the degree of bi-polarization at the outset and the maximal degree of bi-polarization that occurred during the simulation. Following Flache and Mäs (2008a; 2008b), the degree of bi-polarization was measured with the standard deviation of the distribution of pairwise opinion distances between all pairs of agents in the population. This measure takes its maximal value (1) when there are two equally large and maximally different subgroups and. The minimal value of the polarization measure (0) is obtained for perfect opinion consensus. In between these two extremes, the polarization measure increases in the extent to which the opinion distribution is bimodal, with equally large modes at opposite extreme ends of the opinion spectrum. Panel B in figure III.2 shows the average degree of bi-polarization at the beginning of the runs and its average increase. Under all conditions, the simulations started with random, uniform opinion distributions. This resulted for all conditions in a low degree of bi-polarization in the initial situation (indicated by the black areas of the bars). The white areas of the bars show that the average maximal degree of bi-polarization obtained in the simulation runs increased in the strength of homophily, *h*. Furthermore, bi-polarization increased only slightly in the course of a simulation run when homophily was weak (*h*<4). In other words, in these conditions the simulated populations hardly bi-polarized. Only strong homophily could give rise to significant levels of bi-polarization.

Figure III.2: Results from simulation experiment on the effects of homophily on bi-polarization (500 runs per condition, *N*=20, *P*=*C*=20, *S*=6)



Finally, for those runs which ended in bi-polarization, panel C in figure III.2 informs about the average number of simulations events that it took to reach the equilibrium. This measure serves as an indicator of the duration of the bi-polarization process. The conditions with weak homophily (h<3) are neglected in panel C, because only a single run ended in bi-polarization in these conditions. The graph shows that the weaker homophily the more events it took until bi-polarization was reached. These results indicate that bi-polarization is not only possible under strong homophily. The self-reinforcing process that leads to bi-polarization may evolve also if homophily is only moderately strong. However, the graph shows that for the corresponding conditions it may take a considerable amount of time until a group splits up into opposing factions. But the longer it takes before the equilibrium of bi-polarization is reached, the more likely it is that in the process agents interact with dissimilar others and learn counter arguments to their current opinion tendency. If this happens, it is likely that agents further spread these counter arguments within the subset of the population that leans towards the same pole of the opinion spectrum. As a consequence, bi-polarization declines again and the population becomes more likely to move towards the other possible equilibrium, perfect consensus. This explains why under moderate homophily perfect group splits occur only rarely.

To summarize, our computational experiments yield two main findings. First, the informal reasoning that we proposed above is consistent: The interplay of homophily and argument exchange can entail bi-polarization. Second, the social influence dynamics of our model only generate bi-polarization if homophily is sufficiently strong. To assure that these conclusions can be generalized beyond the specific parameter setting that we inspected in the experiments which we report in this paper, we have conducted extensive additional tests, varying the remaining parameters of our model (N, P, C, S). We have not found any combination of these parameters that generated bi-polarization under weak homophily (h<2). This suggests that strong homophily is a *necessary* condition of bi-polarization. Accordingly, the strength of homophily is one of the two major manipulations in the experimental test of our theory of bi-polarization.

III.4. Experimental Test

III.4.1. Overview

The two key mechanisms that according to our theory jointly underlie bi-polarization are argument exchange and homophily. To test this theory, we designed and conducted a controlled group discussion experiment in which both mechanisms were manipulated independently from each other. To assure that bi-polarization in the group discussions can not be attributed to the negative influence mechanism, it was carefully avoided that participants received information which would allow them to perceive in- or outgroups within the set of their interaction partners. The key hypothesis of the experiment was that bi-polarization would nevertheless occur in the group discussions, but only in those conditions in which participants could exchange the arguments underlying their opinions, and in which selection of interaction partners was driven by homophily based on opinion similarity.

The group discussion was designed such that groups of eight participants interacted in a computer network which we implemented with the software z-tree (Fischbacher 2007). An experimental session consisted of seven interaction periods. In each period, participants were matched in pairs and exchanged information concerning an artificial issue. At the beginning of the experiment and after every interaction, we measured the opinions of the participants and calculated the degree of bi-polarization, the dependent variable of this study.

Computer-based communication was chosen to preclude that participants could be negatively influenced by interaction partners which the subject would categorize as members of some subjectively perceived outgroup. Participants could transmit in the communication process only opinion ratings and/or standardized arguments. In this way, participants were not aware of any social characteristic of their interaction partners.

Homophily was manipulated via the procedure with which participants were matched to interaction partners in the course of the group discussion. We compared in a within-subject design two different matching procedures. The entire group discussion process consisted of seven interaction periods. In the first three interaction periods, only participants with similar opinions interacted with each other (homophilious matching). In the remaining four periods, only participants with dissimilar opinions were brought into contact. We expected that – given the possibility of argument exchange – bi-polarization will increase in the homophilious matching phase, while it will decline in the phase of heterophilious interaction.

In the homophilious matching phase (first three interactions), participants are matched with partners who hold similar opinions. According to existing theories of bi-polarization (Baldassarri and Bearman 2007; Hogg, Turner and Davidson 1990; Macy et al.

2003), there should be no systematic tendency for opinion change either towards or away from interaction partners' opinions in this phase. In contrast, our theory of bi-polarization posits that opinions will change when participants can exchange arguments. More specifically, our theory predicts that if there is argument exchange, participants will shift in the homophilious interaction phase their opinions towards the pole of the opinion spectrum towards which they tended already prior to interaction. We have assured in the experimental design that already at the outset arguments are distributed such that there is some tendency for the group to split into two subsets of participants which lean towards opposite poles of the opinion spectrum. Accordingly, our theory further predicts that bi-polarization will increase in the homophilious interaction phase, if argument exchange is possible. Bi-polarization will, however, not increase in the homophilious interaction phase according to our theory if participants can not learn arguments in the course of interaction.

To test these predictions, we compare opinion dynamics in groups where participants discussed the issue only with arguments (*Only-argument-condition*) with groups where participants could only inform each other about their opinions (*Only-opinion-condition*). During the first three interactions of the *Only-argument-condition*, both preconditions of bi-polarization are satisfied. Hence, we expect a significant increase in the degree of bi-polarization during the first interactions in this condition. We expect *no* bi-polarization in the *Only-opinion-condition*, because negative influence tendencies are precluded in the experiment. For the same reason, we predicted bi-polarization during the first three interaction periods (homophilious matching) to be stronger in the *Only-argument-condition* than in the *Only-opinion-condition*.

As an additional test, we included an experimental condition where interaction partners exchanged both arguments and opinions (*Opinions and arguments-condition*). This serves as robustness test of our theory. Two outcomes are plausible. First, one might expect that the effect of learning a new argument on a participant's opinion is reduced, if participants learn that their interaction partners do not hold more extreme opinions than they themselves do. Following this reasoning, one would expect weaker bi-polarization in the *Opinions and arguments-condition* compared to the *Only-argument-condition*. Contrary to this reasoning, information about the interaction partner's opinion could also intensify the effects of persuasive arguments. In the *Opinions and arguments-condition*, participants were made aware of the fact that their first three interaction partners held similar opinions (due to homophilious matching). As a consequence, participants in this condition might have felt more attracted (Byrne 1971) by the first three interaction partners than participants of the *Only-argument-condition*. This could, in turn, let the interaction partner appear more credible and thus increased the persuasiveness of the transmitted argument. This reasoning implies that the strongest bi-polarization should be observed in the homophilious interaction phase in the *opinions and arguments-condition*.

To summarize, our experiment is designed to test the following main hypotheses.

Hypothesis 1: In the homophilious matching phase, there will be more bi-polarization in the *only-argument condition* than in the *only-opinion condition*.

Hypothesis 2: In the heterophilious matching phase, bi-polarization will decrease in the *only-argument condition*, in the *only-opinion condition* and in the *opinions and arguments-condition*.

III.4.2. Method

Participants. Members of a general subject pool at the Department of Sociology at the University of Groningen had been invited to participate in this experiment. This subject pool comprises students and alumni of the two universities in Groningen. Interested students could register for a specific session using an online form (Greiner 2004). We assigned the sessions randomly to the three experimental conditions. Participants received monetary compensation for participation. After excluding problematic sessions (see below), we included data of 65 female and 31 male participants in the analyses (N=96). On average, participants were 23 years old.

Procedure. In each experimental session, we invited 8 participants to a computer laboratory where they sat in separate cubicles. We informed them that they would not be deceived during the experiment and that we had designed the experiment to study the formation of individual opinions in a social setting. Participants were asked to imagine that they were member of a discussion group, talking about the best location for building a new leisure center. This new center could be constructed in one of two hypothetical towns (town A and town B) or at any place in between these two towns. We chose this artificial issue because participants had no previous knowledge about it. This made it possible to impose the arguments that were known to each of the participants. In addition, the best spot for the leisure center can be identified on an interval scale, providing the participants an unambiguous means to inform each other about their opinion. After all participants had confirmed that they had understood the instructions, we started the computer program which ran the experiment. From now on, all instructions and communications took place on the computers screens.

In the first phase of the experiment, each participant received a different set of three arguments. Each argument suggests that either town A or town B is the better place for the new leisure center. For example, one of the pro town A arguments reads: "There is a university in town A. The nearer the leisure center will be build to town A, the more students will be attracted". Altogether there were 6 arguments pro town A and 6 pro town B. Half of the participants received 2 arguments pro town A and one pro town B and the other half received one pro town A and two pro town B. In the following, we therefore refer to those participants who received two pro town A arguments as "A-types" and to the others as "B-types". Whether a specific participant was of type A or B was assigned randomly. Furthermore, all participants who received two pro town A (B) arguments received the same pro town B (A) argument. This was done to assure that the degree of opinion homophily between members of the same type remained approximately constant throughout the homophilious interaction phase. We made the participants explicitly aware of the fact that they had received different sets of arguments. However, we did not tell them how we distributed the arguments. Hence, we expect that the participants were not aware of the two types and thus no social categorization was possible on basis of the initial distribution of arguments.

After the instructions and assignment of initial arguments, participants were asked to rate for each of their arguments how relevant it was for them on a 7-point scale. We included this to force participants to read each argument carefully and to allow us to check later the plausibility of participants' opinion ratings (see below). Subsequently, participants expressed the first time their opinion about the best location for the new leisure center. We used a scale ranging from -50 (*town A*) to +50 (*town B*). Participants could choose any value between the two extremes.

In the second phase, each participant interacted once with each of the seven other participants of the session. In the first three rounds (homophilious interaction phase) interactions took only place between participants who had received the same number of pro town A and pro town B arguments. In the remaining 4 interactions, participants were subsequently matched with the 4 participants of their session who had another number of pro-A and pro-B arguments (heterophilious interaction phase). We used this interaction schedule in all three between-subject conditions. Participants were not aware of the schedule. We only told them that they would interact once with each participant of the experiment. All interactions did really take place. Participants were not deceived.

In the *Only-opinion-condition*, each interaction consisted of two steps. First, the computer informed the participants about their partners' opinion on the best location for the leisure center, showing the partner's most recent opinion rating. Second, all participants rated again where they personally thought the best place for the leisure center was. The interactions in the *Only-argument-condition* consisted of three steps. First, both interaction partners were asked to select which of their arguments should be transmitted to their current interaction partner. Second, participants read which argument their respective partner had transmitted. Whenever a participant had received a new argument then this argument was added to this participant's list of arguments and could later be transmitted to interaction partners. Finally, the participants expressed their opinion again. The new opinion rating, however, was not communicated to the current interaction partners. The *Opinions and arguments-condition* was very similar to the *Only-argument-condition* except for the fact that in step 2, participants did not only read the transmitted argument but also learned the opinion of the respective partner about the best location of the leisure center.

At the end of each interaction, participants read on the screen that the interaction was finished and that they would be matched with a new interaction partner. When all 7 interactions were completed, participants answered a short questionnaire and received monetary compensation for participation.

Altogether, we conducted 18 sessions with 8 participants per session. However, we excluded 6 sessions from the analysis because the manipulation of the initial opinions did not work out. Even though participants received at the beginning of the experiment two arguments favoring one of the two towns, it was still possible that participants considered the one argument in favor of the other town as most relevant. In some cases, participants' initial opinion therefore tended towards the town for which less arguments were given. All sessions in which this happened for more than one of the participant were excluded from the analysis. This was necessary to ensure that the interaction schedule imposed homophilious matching during the first three interactions. Altogether, we used data from twelve sessions with eight participants each for the statistical analyses (N=96). For each of the three conditions, four sessions are available (N=32 each).

Six participants misunderstood the answering scale and entered their opinion with the wrong sign at the very first measurement (e.g. 20 instead of -20). At the second measurement, these participants corrected their opinion to the intended value (-20). We could identify these participants because the relevance measures for the initial arguments where not in line with the initial opinion (the counter argument was rated less relevant than the other two arguments). This misunderstanding and, in particular, the participants' corrections at the second measurement might be problematic for parts of our analyses. The reason is that the initial degree of bi-polarization may be underestimated and thus the change of bi-polarization between the first and the second interaction round be overestimated. To avoid these complications, we reversed the sign of the initial opinion of the six participants. As a result of these changes, the increase in the degree of bi-polarization is unaffected by the participants' misunderstandings. To be sure, these changes are a conservative correction, because they make it more likely that we refute the hypothesis of *increasing* bi-polarization in the homophilious matching phase with argument exchange. As a further control, we present in the Appendix additional analyses that used the absolute value of the participants' opinions as an indicator of polarization. This variable is unaffected by the corrections that we conducted. The additional analyses show that all results could be replicated with this method.

III.4.3. Results

Figure III.3 describes bi-polarization in the three conditions of the experiment. In each graph, the upper (lower) thin solid line depicts the average opinion of the 16 participants of type B (A). The distance between the two lines (highlighted by the gray area) serves as a measure of the opinion distance between the two types. It neglects, however, the degree to which opinions vary between participants of the *same* type. To also take this into account, we calculated for each session and each interaction period the bi-polarization measure that we have also used in the computational experiments. The change of the average value across all sessions of this bi-polarization measure is shown with bold solid lines in figure III.3. For the opinion scale of the experiment the bi-polarization measure can maximally take the value of 50 (maximal degree of bi-polarization). The measure decreases to zero if all participants of a session have exactly the same opinion. The dotted lines in the graphs highlight the changes in average degree of bi-polarization during the first three (homophilious matching) and the last four interactions (heterophilious matching). To also quantify bi-polarization in the three conditions, we estimated for each condition separately a linear regression with the degree of bi-polarization as dependent and two period effects as independent variables. The first period effect models the change in bi-polarization during the first three interactions (Periods are coded: 0 1 2 3 3 3 3 3) and the second quantifies the change during the remaining 4 interactions (Periods are coded: 0 0 0 0 1 2 3 4). The regression coefficients of the two period effects indicate whether there was bi-polarization (positive coefficient); no change (insignificant coefficient); or whether opinion distance between the two types decreased (negative coefficient). For each condition there were 32 observations available (4 sessions × 8 opinion measurements).

Figure III.3: Distance between subgroup averages and Bi-polarization dynamics

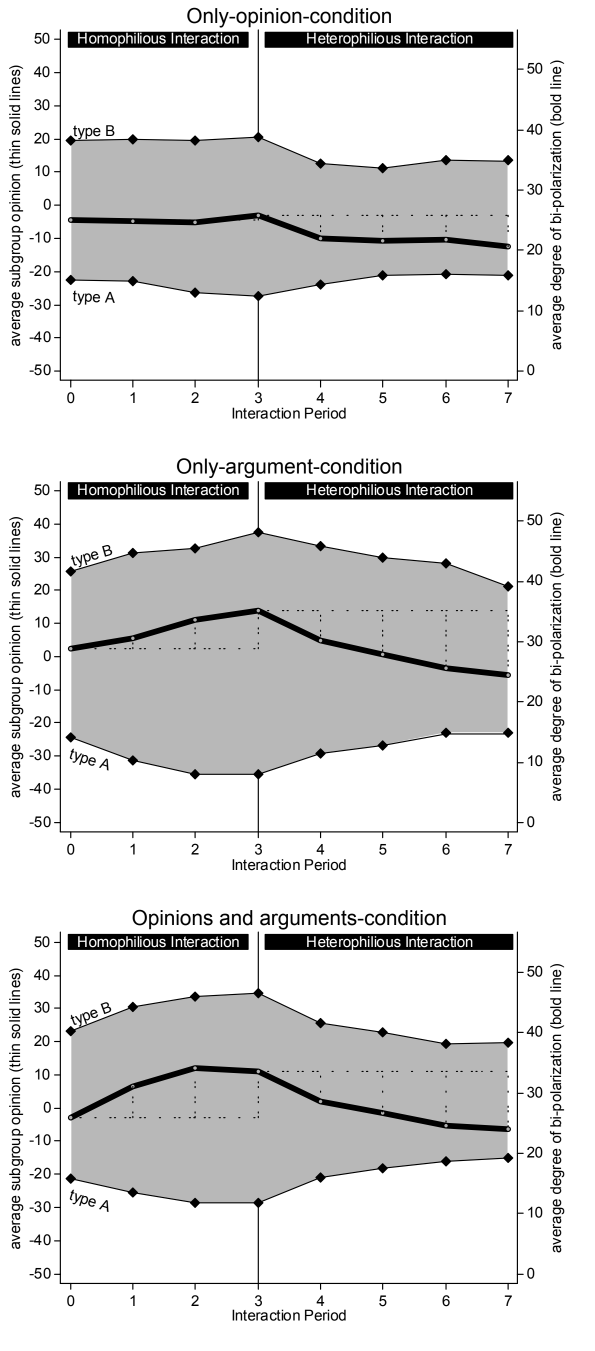


Figure III.3 shows for all three conditions that there have been significant differences between the opinion averages of the two types of participants already before the first interaction (interaction period=0). Also the initial degree of bi-polarization was in all conditions significantly different from zero (t-values of intercepts in the regressions range from 15.03 to 24.43). This demonstrates that the assignment of arguments led to the desired opinion differences between the two types.

In the *Only-opinion-condition*, the degree of bi-polarization hardly changed during the first three interactions. Actually, it *de*creased on average by 0.21 during the first three interactions. The decrease is not significantly different from zero (t=-0.42). This is different for the *Only-argument-condition* and the *Opinions and arguments-condition*. In both conditions the degree of bi-polarization significantly *in*creased per interaction, by 1.71 (t=2.98) and 1.91 (t=2.15) respectively.

Dynamics changed when interaction partners with different opinions were matched (interaction period>3). Under all three conditions, figure III.3 shows decreasing opinion differences between the two types of participants. This confirms our second hypothesis. In the *Only-opinion-condition*, the degree of bi-polarization decreased on average by 1.01. This effect differs significantly from zero (t=-2.77) but the confidence interval of the effect reveals that it does not differ significantly from the weak decrease during the first three interactions. In the *Only-argument-condition* and the *Opinions and arguments-condition*, the degree of bi-polarization decreased from interaction period 4 on by 2.75 (t=-6.5) and 2.74 (t=-4.17) on average. In both conditions, this decrease during the interactions 4 thru 7 differs significantly from zero and therefore also from the *in*crease during the first three interactions.

We also wanted to know whether the dynamics of bi-polarization differed significantly between the three conditions. To test this, we began by estimating a regression that tested differences between the *Only-opinion-condition* (reference category) on the one hand and the *Only-argument-condition* and the *Opinions and arguments-condition* on the other hand. For this purpose, we used the same regression approach as for the separate models and included main and interaction effects for the experimental conditions. It turned out that the increase in the degree of bi-polarization during the first three interactions was in significantly stronger in both the *Only-argument-condition* (t=2.01) and the *Opinions and arguments-condition* (t=2.22) than it was in the *Only-opinion-condition*. This supports our hypothesis 1. We observed bi-polarization even in the absence of negative influence, but only in the conditions in which both homophily and argument exchange were imposed.

We found furthermore that the *de*crease in the degree of bi-polarization during the final four interaction periods was significantly stronger in the *Only-argument-condition* (t=-

2.47) and the *Opinions and arguments-condition* (t=-2.44) as compared to the *Only-opinion-condition*. A comparison of the differences between the *Only-argument-condition* and the *Opinions and arguments-condition* revealed that both the developments during the first three interactions (t=.19) and the subsequent four interactions (t=.02) did not differ significantly between the two conditions.

In conclusion, the results of the experiment confirm our hypotheses. We found significant bi-polarization in both conditions where participants exchanged arguments under homophilious matching, despite the fact that negative influence was precluded by the experimental design (Hypothesis 1). In the control condition, where participants only exchanged opinions, we did not find significant bi-polarization. Thus, the empirical test supports the theoretical results of our computational experiments. Bi-polarization can arise without negative influence, but only if homophily in the selection of interaction partners is sufficiently strong. In addition, we do not find significant differences in bi-polarization between the *Only-argument-condition* and the *Opinions and arguments-condition*. This suggests that the bi-polarization dynamics of persuasive arguments are robust to the effects of opinion exchange. At the same time, opinion exchange alone does not entail bi-polarization, as we found in the *Only-opinion-condition*. Finally, the results support that homophily is a precondition of bi-polarization. Under all three experimental conditions, we find decreasing differences between the two types of participants under heterophilious matching (Hypothesis 2).

III.5. Summary and Discussion

Previous explanations of bi-polarization hinge on the assumption that individuals seek to adopt opinions that maximize opinion differences to outgroup members (e.g. Hogg, Turner and Davidson 1990; Mason, Conrey and Smith 2007). Yet, empirical research which tested the assumption of negative influence has produced very mixed results and is open to methodological criticism (e.g. Krizan and Baron 2007). This has led us to propose a new theory of bi-polarization that does not rely on negative influence. Our theory has two main ingredients. First, it is assumed that individuals tend to interact with others who hold similar opinions (homophily). Second, during interaction individuals influence each others' opinions by exchanging arguments. Based on PAT, we assume that when individuals with similar opinions interact, they likely provide each other with new arguments that support their opinions. The interplay of homophilious selection of interaction partners and influence with persuasive arguments can create a feedback process which triggers the formation of subgroups that independently from each other develop increasingly distant opinions.

We studied the theory first from a theoretical angle and then tested it empirically. For the theoretical study, we developed a formal model of the theory. Computer simulations led to three main results. First, we demonstrated that our formal model can generate bi-polarization, although the model does not include the assumption of negative influence. Second, we showed that the interplay of homophily and argument exchange can trigger bi-polarization even in a population where initially all individuals hold perfectly similar opinions. This implication of our theory contradicts self-categorization theories, which predict bi-polarization only when individuals perceive differences between members of the population. Third, a simulation experiment revealed that according to our theory bi-polarization can only arise if homophily is sufficiently strong.

Future research should use the formal model to identify further conditions of bi-polarization. Mäs et al. (see chapter IV) have recently explored the effects of including demographic characteristics into the formal model (Mäs et al. 2008). Interestingly, their simulations revealed that demographic diversity in a group may intensify bi-polarization only if two further conditions are met. First, the demographic attributes need to form a sufficiently strong faultline (Lau and Murnighan 1998) in the sense that several demographic dimensions are sufficiently correlated. Second, demographic attributes need to be correlated with agents' opinions (Homan et al. 2007; Phillips 2003; Phillips et al. 2004). If one of the conditions is not met, then subgroups fail to become sufficiently distinct from each other. The reason is that similarity between subgroups in demographic characteristics or initial opinions provides a basis for argument exchange accross subgroup boundaries which, in turn, prevents bi-polarization. In addition, simulation results suggest that demographic faultline may fire up bi-polarization in the short run, but at the same time they help to overcome opinion differences in the long run as long as there is at least some demographic overlap between subgroups (Mäs et al. 2008).

In the second part of the present paper, we tested our theory with a laboratory experiment. We created a setting for which existing theories of bi-polarization assume no negative influence and therefore predict no bi-polarization. Yet, we predicted and found bi-polarization for specific experimental conditions. In particular, we found significant bi-polarization under the conditions that our formal model identified, strong homophily and the possibility to exchange arguments.

We have deliberately created an artificial setting in our experiment. This allowed us to exclude other explanations of bi-polarization than the mechanisms that we wanted to test. Yet, we believe that it would be fruitful for future experimental research to test our theory of bi-polarization in less constrained settings, like standard group discussion experiments (Johnson and Johnson 1982). At the same time, we also believe that the setting of our experiment may be a realistic model of some parts of the social world. For instance, internet search engines and online social networks make it very easy to connect only to others with similar opinions (Sunstein 2008) and exchange very condensed pieces of information, like in our experiment. We therefore believe that future empirical research may gain a better understanding of opinion dynamics in real life situations by focusing on those settings in which both argument exchange and homophily are possible and could thus, according to our theory, give rise to bi-polarization.

III.6. Appendix to chapter III

In laboratory experiment, it turned out that six participants misunderstood the answering scale and entered their opinion with the wrong sign at the very first measurement (e.g. 20 instead of -20). At the second measurement, they corrected their opinion and entered the intended value (-20). We had to rectify this because otherwise the statistical model would spuriously indicate that the degree of bi-polarization increased. Above, we have argued that these changes make it more likely that our hypothesis is refuted. Still, we provide here additional analyses which demonstrate that our results are not affected by the changes.

In particular, we have focused on the participants level and used the absolute value of each participant's opinions as the dependent variable, a measure which is unaffected by the corrections. According to the hypotheses, we expect that in the conditions where arguments were exchanged the absolute value of the participants' opinions increased during the first three interaction periods. This reflects that the participants developed more extreme opinions. In contrast, we expect that the opinions remained constant in the *Only-opinion-condition*.

To take into account that the opinion measurements of each participant are interdependent, we estimated linear multi-level regressions (Snijders and Bosker 1999) with random intercept effects on the participant level. An empty model revealed that the explained variance on the group level was not significant. Including random period effects (random slopes) did not affect the significance decisions.

We followed the same strategy as with the group-level analyses in the main paper. Models 1 thru 3 (see table III.1) tested the two period effects separately for each condition. In line with the analyses of the group-level data, we found no significant opinion extremization during homophilious matching in the *Only-opinion-condition* (model 1). In contrast, we found significantly more extreme opinions in the *Only-argument-condition* and the *Opinions and arguments-condition* (models 2 and 3). Also in line with the group-level analyses, we found for all conditions that the participants held increasingly moderate opinions during heterophilious matching of interaction partners.

Table III.1: Multi-level regression of participants' absolute opinion value on interaction period

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5
I: only opinions					
Intercept	25.89 **			25.89 **	
	(2.20)			(2.28)	
Homophil. interaction	-.17			-.17	
	(.34)			(.46)	
Heterophil. interaction	-1.41 **			-1.41 **	
	(.25)			(.34)	
II: only arguments					
Intercept[a]		30.48 **		4.59	30.48 **
		(2.33)		(3.22)	(2.31)
Homophil. Interaction[b]		1.89 **		2.07 **	1.89 **
		(.49)		(.65)	(.51)
Heterophil. Interaction[b]		-2.95 **		-1.54 **	-2.95 **
		(.36)		(.48)	(.37)
III: arguments and opinions					
Intercept[a,c]			29.19 **	3.30	-1.29
			(2.29)	(3.22)	(3.27)
Homophil. Interaction[b,d]			2.11 **	2.28 **	.22
			(.52)	(.65)	(.72)
Heterophil. Interaction[b,d]			-3.66 **	-2.25 **	-.70
			(.39)	(.48)	(.53)
Random Effects					
Sd(intercept)	11.80	11.86	11.42	11.70	11.65
	(1.53)	(1.57)	(1.53)	(.89)	(1.10)
Sd(resid)	4.80	6.91	7.43	6.49	7.18
	(.23)	(.33)	(.35)	(.18)	(.24)
Number of participants	32	32	32	96	64
Number of observations	256	256	256	768	512

Note. Table reports unstandardized effects. Values enclosed in parentheses represent standard errors.
** significant on 0.01 level
[a] in model 4 included as condition dummy. [b]in model 4 included as interaction term with condition dummy. [c] in model 5 included as condition dummy. [d]in model 5 included as interaction term with condition dummy.

Model 4 compares the *Only-opinion-condition* with the *Only-argument-condition* and the *Opinions and arguments-condition*. In line with the group-level results, we find that the opinions turned significantly more extreme during homophilious matching in the two conditions where arguments were exchanged. Also the opposite effect during heterophilious matching is stronger in the two conditions with argument exchange, compared to the *Only-opinion-condition*.

Finally, model 5 compares the *Only-argument-condition* and the *Opinions and arguments-condition*. Supporting the group-level results, table III.1 shows that there are no significant differences between the two conditions.

We replicated all five models and included the respondents' age and gender as control variables. In all five models, neither the main nor interaction effects with the period effects and the between-subject conditions were significant.

IV. Argument Exchange and Demographic Faultlines[14]

Abstract
In the previous chapter, we presented a new theory of opinion polarization. In this chapter, we will apply this model to the research on demographic faultlines in work teams. We will test if the new model also predicts the polarizing effects of demographic faultlines that we have identified in the negative-influence model (see Chapter II).
We will show that the argument-exchange model supports the faultline hypothesis. However, the new model points out several conditions for this effect which previous contributions have overlooked. First, even with a very strong faultline, opinions will only polarize in groups where individuals tend to select similar interaction partners. Second, polarization is more likely the stronger opinions and demographic attributes in a team are correlated initially, that is, prior to interaction between the group members.
Furthermore, our new model implies that the short term effects of demographic faultlines differ crucially from their long term effects. Groups where demographic attributes are not perfectly correlated will eventually arrive at consensus even though they might suffer from polarization in the short run. Counter-intuitively, the model implies that the convergence process is faster the *stronger* the demographic faultline is.

IV.1. Introduction

Demographic diversity at the workplace is a major challenge for organizations and is becoming an increasingly important issue as the economy globalizes (for comprehensive reviews about theoretical and empirical research see: Bowers, Pharmer and Salas 2000; Milliken and Martins 1996; Pelled 1996; Stewart 2006; van Knippenberg and Schippers 2007; Webber and Donahue 2001; Williams and O'Reilly 1998). For work teams, demographic diversity can be beneficial, because it broadens the social and human capital

[14] This chapter was written together with Andreas Flache, Károly Takács and Karen A. Jehn and is currently under review. The title of the original article is: "In the short term we divide, in the long term we unite. Crisscrossing work team members and the effects of faultlines on subgroup polarization"

of the team. However the benefits do not accrue automatically. Demographic dissimilarity between team members may, at the same time, cause conflicts and tensions and thus threaten performance. This lead Milliken and Martins to conclude in their review of the field that "diversity thus appears to be a double-edged sword"(1996: 403).

In the search for conditions that explain why diversity sometimes increases team performance and reduces it at other times, Lau and Murnighan (1998; 2005) proposed that the effects of diversity may decisively depend on the way demographic attributes, like age and gender, are distributed among team members. Their main hypothesis is that diversity impairs team functioning when the distribution of demographic attributes generates a *strong faultline*. "Group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of resulting subgroups. In contrast, faultlines are weakest when attributes are not aligned and multiple subgroups can form" (Lau and Murnighan 1998: 328). They argue that diversity (demographics not aligned) increases the potential of a team for creativity and good performance but when the diversity is in a group with a strong faultline (demographics aligned), this potential may not be realized. The team may split up into subgroups with polarized opinions that cause conflicts between team members (Bezrukova, Thatcher and Jehn 2007). An intriguing implication of this theory is that an ideal workgroup composition might exist such that large pools of social and human capital can be obtained, but the damaging effects of diversity on cohesion can be avoided.

We contribute to the faultline research by elaborating the explanation of the faultline hypothesis (Lau and Murnighan 1998) and thereby revealing crucial implications of faultline theory that have been overlooked so far. We start by reviewing two theories that have been used to explain faultline effects and that are based on fundamentally different arguments. The first, Lau and Murnighan's (1998) theory, highlights that the interplay of homophilious selection of interaction partners with social influence breeds subgroup polarization in work teams with strong faultlines. Subgroup polarization is our main dependent variable and is defined as the degree to which a work team separates into subgroups that hold opposing opinions on work-related issues. The second theory has been developed almost a century ago in the classical sociological and anthropological literature and focuses on the integrating function of "crisscrossing actors" (Colson 1954; Evans-Pritchard 1939; Flap 1988; Galtung 1966; Lijphart 1977; Nieuwbeerta and Flap 2000; Ross 1920; Simmel 1922 (1908)). These actors share at least one demographic attribute with another demographic subgroup than their own and thus, function as a bridge over the subgroup split that was caused by the faultline.

Lau and Murnighan (1998) did not take into account how crisscrossing actors can reduce subgroup polarization when the faultline is strong. We show that while their theory implicitly considers this, by not examining crisscrossing actors, the authors failed to realize some crucial implications of their theory. Foremost, we show that their theory predicts subgroup polarization only in the *short* term. We propose that in the long run, crisscrossing actors help to overcome subgroup polarization and group splits even in teams with very strong faultlines. Moreover, we show that subgroup polarization in the short term depends upon two further conditions that Lau and Murnighan implicitly assumed, but which they did not examine theoretically. First, strong faultlines entail subgroup polarization in the short run only when employees exhibit sufficiently *strong homophily* when selecting partners in their communication with other team members. That is, team members have a sufficiently strong preference for interacting with colleagues that are similar to them on certain attributes (McPherson, Smith-Lovin and Cook 2001). Second, we propose that faultlines entail group splits in the short run only when there is sufficient initial congruency between work-related opinions and demographic attributes in a work team (Homan et al. 2007: 82; Phillips 2003: 7; Phillips et al. 2004: 503).

Faultline effects result from a complex interplay of the interactions between multiple team members. As Harrison et al. (2007) have suggested, agent-based computational modeling is a powerful research method that allows to cope with this theoretical complexity and to reveal counter-intuitive implications of a theory. Frank and Fahrbach (1999) developed an agent-based model of complex and interrelated network and opinion dynamics in organizations, based on mechanisms that are very similar to those we assume. We follow their lead and use a formal modeling approach to show how Lau and Murnighan's reasoning, on the one hand, and the theory of crisscrossing actors, on the other hand, can be reconciled. We present and analyze a computational model that is based on Lau and Murnighan's assumptions and test if our new propositions do consistently follow from their theory. Our results support the proposition that faultline effects occur only in the short run and that strong homophily and initial congruency are crucial conditions for the effect of faultlines on group polarization. Finally, our formal analyses reveal a counter-intuitive effect of faultline strength. It turned out that the same communication structures that trigger short-term subgroup polarization in teams with strong faultlines accelerate the process of consensus formation in the long run. Teams with strong faultlines might arrive at a consensus faster than teams with weak faultlines.

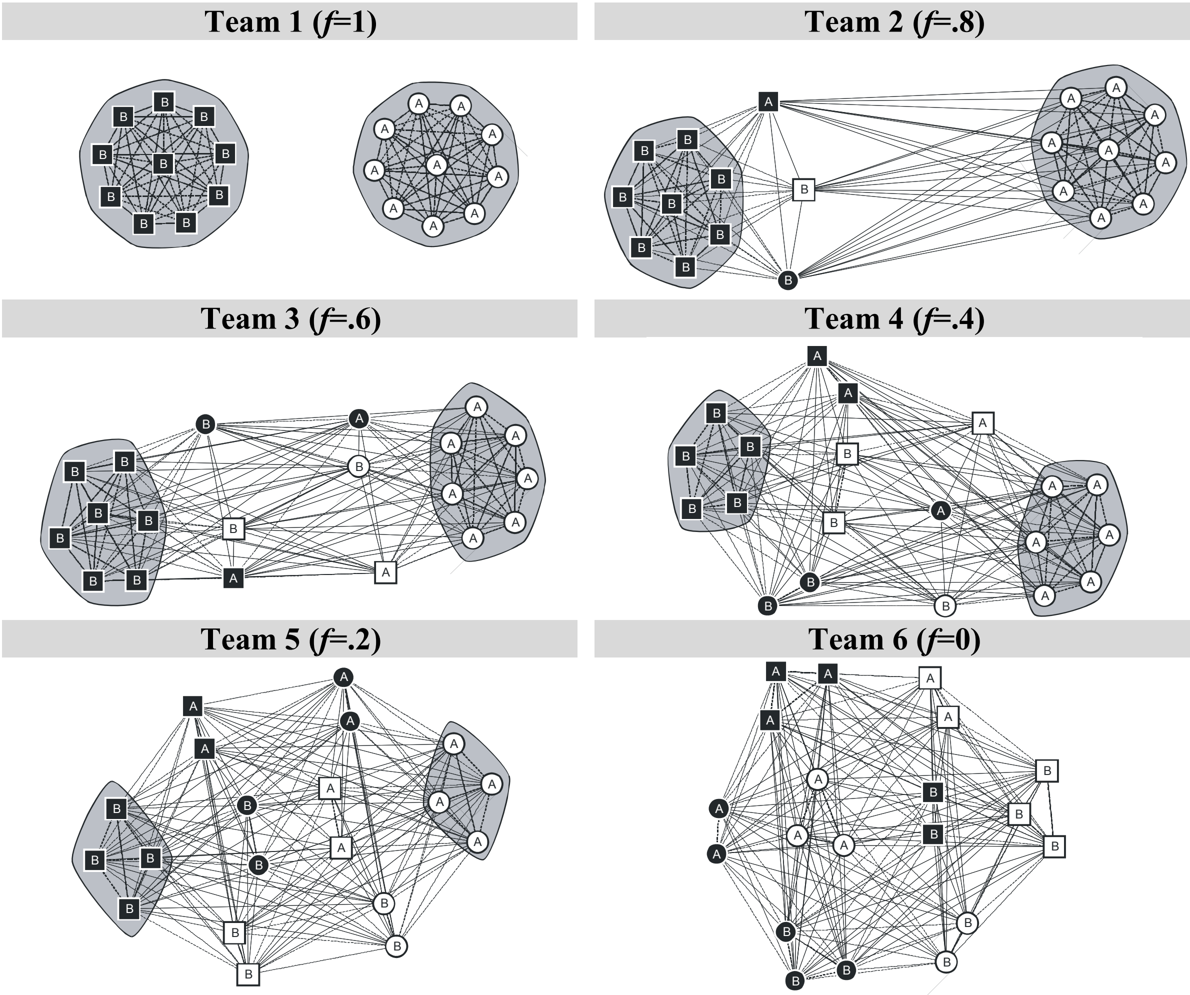
IV.2. Two Explanations of faultline effects

IV.2.1. Lau and Murnighan's explanation of faultline effects

Lau and Murnighan argued that all newly formed teams go through a "sensemaking process of understanding each other and their task" (1998: 332) to coordinate similar opinions about what their task is, how to fulfill it, and how to devise work. In this process, the interplay of two core mechanisms can cause problems in teams with a strong faultline. First, Lau and Murnighan assume *homophilious selection of interaction partners*. Team members tend to associate with colleagues who share relevant demographic attributes. This assumption is prominently supported by a large body of sociological research on homophily (Lazarsfeld and Merton 1954) or the tendency of "birds of a feather flock together" that has consistently been identified as a strong force in social interactions (McPherson, Smith-Lovin and Cook 2001). Studies in both educational (e.g. Kandel 1978; Moody 2001) and organizational settings (e.g. Bacharach, Bamberger and Vashdi 2005; Ibarra 1992; McPherson and Smith-Lovin 1987; Ruef, Aldrich and Carter 2003) have provided empirical confirmation of the homophily concept.

Homophilious selection of interaction partners implies that the communication structures in a team crucially depend on faultline strength. To show this, we constructed 6 hypothetical teams of 20 individuals (see Figure IV.1). This size is not too big to be unrealistic for a work team, but also large enough to allow for a sufficiently fine-grained variation in the strength of demographic faultlines. Each team member is described by three dichotomous demographic attributes (symbolized as *black* vs. *white*; *A* vs. *B* and *rectangle* vs. *circle*). Diversity in all teams and on all demographic dimensions is kept at its maximum. That is, all three dichotomous variables have a distribution in which both values of the attribute are equally frequent (50:50). Teams differ, however, in the strength of the demographic faultline. Faultline strength is denoted by the symbol f and is measured here by the pairwise Pearson-correlation between all pairs of demographic attributes. Applying a method proposed by Flache and Mäs (2008a; 2008b) we varied faultline strength between maximal (all pairwise Pearson correlations are 1) and minimal faultline strength (all Pearson correlations are 0).

Figure IV.1: 6 hypothetical teams with different faultline strength



The network pictures shown in Figure IV.1 depict how faultline strength and the homophily mechanism shape the communication structure in the 6 work teams. Team members are represented by nodes. A pair of nodes is connected with a line if they have at least one demographic attribute in common. To depict the effects of homophilious selection of interaction partners, nodes have been arranged such that two individuals are nearer to each other the more attributes they share (Kamada and Kawai 1989; McFarland and Bender-deMoll 2007). The dashed circles identify the biggest subgroups of maximally similar individuals. In the team with the strongest faultline (team 1) the three demographic attributes correlate perfectly. Each pair of actors is either maximally similar or maximally *dis*similar and therefore either interacts frequently or never. On the team level, Figure IV.1 depicts two perfectly homogeneous but unconnected subgroups for team 1. As faultline strength decreases, however, this separation between subgroups becomes weaker and completely disappears as faultline strength is minimal (team 6). For instance, there are still two clearly distinct subgroups in team 3 (medium faultline strength). However, the subgroups are smaller and there are also team members that can not be categorized into one of the subgroups. These actors share demographic attributes and therefore also interact with both subgroups. Note that these communication structures are a logical implication of faultline strength and the homophily mechanisms.

In addition to homophily, Lau and Murnighan assume that during interaction, team members exert influence on each others' opinions by *exchanging persuasive arguments* (Isenberg 1986; Myers 1982; Myers and Lamm 1976; Vinokur and Burnstein 1978). "Group members who support similar attitudinal positions will find that, as other members support that position using arguments different from their own, they each have more reason to become even more extreme than they were before" (1998: 332). Research on "polarization" (Myers 1982) has demonstrated how group members tend to become more extreme during group decision making. Faultline theory examines, however, not just polarization within one group, but it focuses on what we denote "subgroup polarization", the degree to which a work team separates into subgroups *holding opposing opinions* (Lau and Murnighan 1998). Subgroup polarization during a team's sense making process is problematic because it breeds emotional conflicts between the subgroups (Lau and Murnighan 1998) which, in turn, hamper good team performance (Jehn 1994).

The interplay of homophilious selection of interaction partners and influence with persuasive arguments can lead to subgroup polarization in groups with strong faultlines. As shown in Figure IV.1, homophily creates subgroups in teams with strong faultlines. Within subgroups, team members frequently exchange arguments but argument exchange *between* subgroups is rare. Lau and Murnighan argue that under these conditions, small initial opinion differences between the subgroups might be amplified during the sense making process. This is because subgroup members will mostly hear and share arguments that support their initial opinions (Stasser 1988), causing opinions in both subgroups to shift towards opposing ends of the opinion scale simultaneously. In other words, subgroup polarization increases. By contrast, in teams with weak faultlines, group members interact with colleagues who hold a variety of different opinions, such that no self-reinforcing dynamic towards emergent subgroup polarization can develop. This mechanism implies the central proposition of Lau and Murnighan's theory.

Proposition 1: The stronger the faultline in a work team is, the stronger subgroup polarization will be.

IV.2.2. The sociological explanation of faultline effects

Almost a century ago classical sociological and anthropological research on social order in stateless societies (Colson 1954; Evans-Pritchard 1939; Flap 1988; Galtung 1966; Lijphart 1977; Ross 1920; Simmel 1922 (1908)) revealed that strong faultlines may cause a problem for social integration. For instance, Ross argued in 1920 in a textbook:

> "Suppose that at a given moment there is a certain strain along the line between Christians and Jews. If now, a strain appears along a quite different line, e.g. that between employers and workman, the religious opposition will become less intense. For Jewish bosses and Jewish workers will be estranged, likewise Christian bosses and Christian workman. On the other hand, Jewish and Christian capitalists will recognize that they are 'in the same boat', while Jewish and Christian workers will sympathize with one another as fellow victims of exploitation. Take the case of a tension between blacks and whites. If the lines of cleavage cross, each opposition will weaken the other. But if, as sometimes happens, all the employers are white and all the employed are black men, then one antagonism reinforces the other and the rift in society is deeper then ever. So, paradoxical as it may sound, a society riven by a dozen oppositions along lines running in various directions may actually be in less danger of early break-up than one split along just one line. For each new cleavage narrows the cross cleft, indeed, you might say that the society is sewed together by its inner conflicts"(Ross 1920: 164-165)

Although both faultline theory and the classical sociological and anthropological literature agree on the prediction that strong faultlines breed conflicts, they base their prediction on different sets of assumptions. Whereas faultline theory argues that the interplay of homophily and social influence may result in subgroup polarization, the sociological theory focuses on the integrating function of "crisscrossing" actors. Crisscrossing actors are individuals that share at least one demographic attribute with members of more than one demographic subgroup. Due to demographic similarity, they are attached to members of more subgroups and are thus able to conciliate in case of conflicts. Colson (1954), for instance, pointed to this in her studies of the African Tonga society in the 1940s. Each Tonga identified himself as a resident of his village and as a descendent of one of the Tonga-clans. However, societal rules prevented these two attributes from aligning. First, marriage between members of the same tribe was prohibited. Secondly, it was the man's privilege to choose the village where his family lived. Third, clan membership was organized matrilineally. That is, individuals belonged to their mother's descent group. Consequently each Tonga was attached by kinship to people living in different villages and at the same time by residence and his family members to descendents of different clans. Colson (1954) describes a conflict between the members of two clans that emerged after a member of one clan killed a member of the other clan. The conflict could be settled by persons that belonged to the victim's clan but lived in a village together with members of the murderer's family. Because of their close relationships to both parties they could negotiate between the two groups and the conflict could be resolved.

From this sociological perspective, the faultline hypothesis follows because the more crisscrossing actors there are in a group, the stronger are the integrating forces that prevent conflicts. The number of crisscrossing actors in a group is, in turn, logically related to faultline strength. Figure IV.1 shows that the higher the number of those team members that are not part of one of the subgroups (i.e. crisscrossing actors), the weaker is the faultline. Teams with the maximal faultline strength (Team 1) consist of only two kinds of actors (*black*, *B*, *rectangles* and *white*, *A*, *circles*). There are no crisscrossing actors in this team. The number of crisscrossing actors, however, increases as faultlines become weaker. Team 2, for instance, still consists of two large subgroups. However, there are also three crisscrossing actors present.

IV.2.3. Integrating the two theories: Why time matters

The processes that the two explanations of the faultline hypothesis propose appear to be fundamentally different. On the one hand, Lau and Murnighan (1998) argue that in teams with strong faultlines, subgroups form that develop increasingly different opinions that tear the team apart. The sociological theory, on the other hand, points to those actors that connect the subgroups and prevent conflicts. We argue that it is of great importance for our understanding of faultline effects to analyze how exactly these two processes are related to each other. We have shown that only groups with maximally strong faultlines have no crisscrossing actors. If crisscrossing actors can prevent group splits, does this then imply that their presence might neutralize the mechanism that Lau and Murnighan have described? Or would homophilious selection and persuasive arguments undermine the integrating effects of crisscrossing actors if the faultline is sufficiently strong?

It turns out that the same mechanisms from which Lau and Murnighan derive their faultline hypothesis can also be used to model the effects of crisscrossing actors. But, as we will show, when we explicitly integrate crisscrossing actors into Lau and Murnighan's reasoning, new consequences arise for the effect of faultline strength on the dynamics of subgroup polarization. With our integrating model, we can identify heretofore overlooked conditions under which the integrating effects of crisscrossing actors can be expected to prevail upon the dividing effect of a strong faultline.

Particularly, we argue that in teams with strong faultlines the processes that Lau and Murnighan's describe breed polarization only in the early stage of the sensemaking process.

Later, however, crisscrossing actors will help overcome group splits. Homophilious selection implies that crisscrossing actors interact with members of both subgroups, because they have some demographic similarity with members of each group. Based on persuasive argument theory (Myers 1982), we can expect that they will get arguments from all sides and will also communicate them to all subgroups they interact with. In this way, crisscrossing actors establish indirect communication between the subgroups who fail to interact directly. This may result in a gradual convergence of the subgroups' argument pools and also of their opinions, eventually reaching overall consensus in the work team.

This resonates with formal theories of the dynamics of social influence (Abelson 1964; Anderson 1991; French 1956; Harary 1959) This literature has suggested that social groups inevitably tend to reach consensus on initially controversial opinions, as long as the group has a connected interaction network in which no member is entirely cut off from influence by other group members or external sources. Based on Lau and Murnighan's assumption that demographic overlap implies interaction, crisscrossing actors can be seen as the link that integrates all group members into the network of mutual social influences. This suggests that, in principle, one single crisscrossing actor might suffice to create enough indirect communication between two subgroups to bring their opinions together. Thus, even in a group with a strong faultline, a small number of crisscrossing actors may ensure that no subgroup is entirely disconnected from outside influences. Accordingly, there should be *no* long run effect of faultline strength on subgroup polarization, except for the extreme case of a maximally strong faultline that divides the team into perfectly distinct subgroups. This absence of an effect of faultlines across almost the entire spectrum of possible teams is clearly contrary to what Lau and Murnighan (1998) suggest.

We propose that time is the critical factor and that crisscrossing actors help overcome group splits in the long run. But we also argue that in teams with a strong faultline, the polarizing dynamic of emergent subgroup splits will, in the short run, be stronger than the integrative dynamic of indirect interaction through crisscrossing actors. The reason is that in a team with a strong faultline, the members of the subgroups are by definition exposed to more other group members who are not connected to the outgroup than they are exposed to influences from crisscrossing actors. Accordingly, it is likely that consensus within the subgroups quickly develops and – based on the persuasive argument mechanism – subgroups initially polarize. At the same time, every member of the subgroup still has a positive probability of interacting with a crisscrossing actor at least from time to time. Whenever this happens, there is a chance that an argument from the outgroup is adopted by ingroup members. This argument can subsequently rapidly spread in the ingroup. Due to the homophily principle, ingroup members are highly likely to interact with each other because they have both a high level of consensus and they are demographically similar. In other words, we propose that the same mechanisms that according to Lau and Murnighan imply subgroup polarization in the short term, also imply that subgroup splits are not stable in the long run if the group comprises crisscrossing members.

Proposition 2: Subgroup polarization occurs only in the short run. In the long run, all teams where faultlines are not maximally strong will develop consensus and will overcome subgroup polarization.

IV.2.4. Conditions of the short term effects of strong faultlines

According to propositions 1 and 2, we expect that teams with strong faultlines will polarize in the short term, but will overcome the split in the long run. Moreover, we propose that the processes, that Lau and Murnighan describe, imply short-term polarization only if two necessary conditions are met. Following Flache and Mäs (2008a; 2008b), we argue that under Lau and Murnighan's assumptions the process of subgroup polarization crucially hinges on the assumption that initial congruency is sufficiently strong, meaning that opinions and demographic attributes in a team are already correlated initially, prior to interaction between the team members. If demographically similar group members do *not* share opinions more with each other than they do with demographically dissimilar others, then the exchange of arguments within demographic subgroups will not increase opinion differences between the groups because members of one subgroup do not learn more new arguments pro or con the original opinion than the actors in the other subgroup. As a consequence, subgroup polarization will not occur. Thus, an initial correlation between demographic attributes and opinions appears to be an essential condition for subgroup polarization in work teams.

Proposition 3: Subgroup polarization increases in the beginning of the team dynamics only if the initial congruency is sufficiently high. Even teams with a strong faultline will not polarize if congruency is weak.

We furthermore propose that subgroup polarization can only take place if homophily is sufficiently strong. We define the strength of homophily as the *degree* to which interaction between similar actors is more likely than interaction between dissimilar actors. The strength of homophily in teams might be determined by the institutional context of work teams. For instance, in teams with high task interdependence workers are forced to collaborate with both similar and dissimilar colleagues to fulfill their tasks. Thus, in these teams similarity will only weakly influence the choice of interaction partners. As a consequence, team members interact frequently with members who hold different opinions and will thus be influenced by them. Such a context would make it unlikely that the teams' opinions polarize, even if faultlines are strong (Molleman 2005).

Proposition 4: Subgroup polarization increases in the beginning of the team dynamics only if homophily is sufficiently strong. Even teams with a strong faultline will not polarize if homophily is sufficiently weak.

IV.3. The Model

The exact logical implications of the combination of homophilious selection, persuasive influence, and faultline strength result from a complex interplay of these mechanisms simultaneously operating in multiple actors responding to each others' behavior. The method of computational agent-based modeling (Adner et al. 2009; Harrison et al. 2007; Macy and Willer 2002) provides multiple examples how in such a complex multi-agent system simple theoretical assumptions may generate counter-intuitive implications that would have been overlooked without model formalization (e.g. Frank and Fahrbach 1999; Siggelkow and Rivkin 2006; Stasser 1988). Accordingly, we conducted a strict test of the logical consistency of our reasoning (Adner et al. 2009) with a formal computational model. In fact, we test if the four propositions really follow logically from the assumptions of Lau and Murnighan's theory. Furthermore, our formal analyses revealed a new and unexpected effect of strong faultlines in work teams that has been overlooked in previous theorizing.

Our formal model is based on the two mechanisms of Lau and Murnighan's (1998) informal reasoning: homophilious selection of interaction partners and influence with persuasive arguments. In this model, each of the N team members is represented as an agent i characterized by D *demographic attributes* ($c_{i,d}$) and K *opinions* ($o_{i,k}$), where d refers to the d'th demographic dimension and k to the k'th opinion. The demographic attributes can either take the value 1 or -1 ($c_{i,d} \in \{-1;1\}$) and remain unchanged in the progress of interaction. The opinions of the actors vary between -1 and +1 ($-1 \leq o_{i,k} \leq +1$) and are open to influence.

Agents base their opinions on arguments $a_{k,l}$. For simplicity, we represent arguments as being either in favor of or against holding a pro-opinion ($o_{i,k} > 0$) on the corresponding issue ($a_{k,l} \in \{-1;1\}$). For each issue k there exist P pro arguments ($a_{k,l} = 1$) and C con arguments ($a_{k,l}$ = -1). Which arguments exist in a given work team setting is summarized in the arguments matrix. This matrix has K columns and $P+C$ rows. Cells with a row number smaller than P+1 hold pro arguments, i.e. $a_{k,l}$ = +1. The remaining columns hold con arguments, i.e. a $a_{k,l}$ = -1. Matrix (a) in Figure IV.2 is an example of an argument matrix with one column (k=1) and 3 pro and 3 con arguments per issue (P=C=3).

Figure IV.2: Example of the updating process

(a)

$$\text{arguments} = \begin{pmatrix} +1 \\ +1 \\ +1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

(b) Initial relevancy matrix of $i*$

$$\text{relevancy}_{i*} = \begin{pmatrix} 0 \\ 0 \\ 3 \\ \boxed{2} \\ 0 \\ 1 \end{pmatrix}$$

(c) Initial relevancy matrix of $j*$

$$\text{relevancy}_{j*} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \boxed{3} \\ 0 \\ 2 \end{pmatrix}$$

(d) Updated relevancy matrix of $i*$ (example 1)

$$\text{relevancy}_{i*} = \begin{pmatrix} 3 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

(e) Updated relevancy matrix of $i*$ (example 2)

$$\text{relevancy}_{i*} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ \boxed{3} \\ 0 \\ 1 \end{pmatrix}$$

To take into account the limited cognitive capacities of humans, we assume that agents base their opinion not on all existing arguments but on a sample of S ($S \leq P+C$) arguments. Technically, an agent's opinion on issue k is the average value ($a_{k,l}$) of the arguments the agent considers as relevant (see equation 1). Thus, the more pro (con) arguments an agent's sample of arguments comprises the higher (lower) the value of the agent's opinion will be.

$$o_{i,k} = \frac{1}{S}\sum_{l=1}^{S} a_{k,l} \qquad (1)$$

Furthermore, agents rank relevant arguments that address issue k according to their recency. As we show below, the more recent an argument is, the longer an agent will consider this argument for opinion formation. However, as equation (1) shows, the recency of an argument has no effect on the extent to which an argument shapes the current opinion. We denote recency ($s_{k,l,i}$) for agent i of the argument l that addresses issue k with integer values between 0 and S $s_{k,l,i} \in \{0,...S\}$. A recency value of $s_{k,l,i}$ = 0 indicates that the argument $a_{k,l}$ is *not* relevant for actor i. Values above zero indicate that this argument affects actor i's opinion on issue k. The most recent argument has the recency value of $s_{k,l,i}$ = S, the second most recent argument has the value S-1, and so on. Thus, if three arguments are relevant (S=3), then one has a recency of 1, one has a recency of 2, and one has a recency of 3. See matrix b in Figure IV.2 for an agent's relevance matrix for one issue (k=1), 6 existing arguments per issue (P=C=3) of which only three are relevant at a time (S=3). This agent considers one pro argument and two con arguments. According to equation (1) the agent adopts an opinion of $o_{i,k}$=-1/3.

We model the sense making process of a team as a sequence of events, each event corresponding to one interaction between two agents. An interaction starts with the partner selection phase and is continued by the social influence phase. In the partner selection phase, two agents are matched for interaction, based on homophilious selection. Subsequently, an opinion of one of the interacting agents is updated, based on the persuasive argument mechanism.

IV.3.1. Homophilious selection

We implement the partner selection phase as follows. In each event the computer program first randomly picks an agent i^*. All agents have at all events the same probability to be picked. Then an interaction partner j ($j \neq i^*$) is selected. To incorporate homophily, the probability that actor j is chosen as interaction partner depends on the similarity between i^* and j. As confirmed by empirical research we assume that both demographic similarity (Ibarra 1992; McPherson and Smith-Lovin 1987; McPherson, Smith-Lovin and Cook 2001; Ruef, Aldrich and Carter 2003) and opinion similarity (Byrne 1971) increase the probability to interact. Similarity sim_{i^*j} varies between 0 and 1. A similarity of zero means that the two actors are maximally dissimilar whereas a value of 1 indicates that both hold the same opinions and the same demographic attributes. We assume that all attributes are equally weighted in the calculation of similarity. Formally,

$$sim_{i^*,j} = \frac{1}{2 \cdot (D+K)} \left(\sum_{d=1}^{D} 2 - \left| c_{i,d} - c_{j,d} \right| + \sum_{k=1}^{K} 2 - \left| o_{i,k} - o_{j,k} \right| \right) \tag{2}$$

The probability that actor j is selected as interaction partner (p_j) is derived from the relative similarity of i^* and j compared to the similarities of i^* to all other actors, except i^* herself. To vary the *strength of homophily* we include furthermore a parameter h into the model (h>0). The higher the value of h, the more the relative similarity of the focal agent i^* and agent j increases the likelihood that j will be chosen as an interaction partner. Technically,

$$p_j = \frac{\left(sim_{i^*j}\right)^h}{\sum_{j=1, j \neq i^*}^{N} \left(sim_{i^*j}\right)^h} \tag{3}$$

Equation 3 shows that the more similar j is to i^*, the higher is the probability that they will interact. If two actors differ maximally with regard to their opinions and their demographic attributes then the probability of interaction equals zero.

IV.3.2. Persuasive Arguments

After the interaction partners i^* and j^* have been selected for the respective event, agent i^* is influenced by j^*. We implement social influence through persuasive arguments in two steps. First, one of the arguments that j^* considers as relevant is adopted by i^*. For this, one of the K opinions is selected randomly for update (k^*), with the same probability ($1/K$) for all opinions. Then one of the S arguments that are relevant for j is picked (a_{k^*,l^*}) with equal probability ($1/S$) for all relevant arguments. Arguments that are not relevant for j^* are not chosen. The chosen argument is adopted by i^*. Technically, the argument a_{k^*,l^*} in i^*'s relevance matrix adopts the value $S+1$ ($s_{k^*,l^*,i^*} = S+1$).

When an agent's relevancy matrix has been updated repeatedly, it is likely that all existing arguments have been adopted at least once. However, it does not seem reasonable that after some time agents consider all arguments as relevant. To avoid this, we implemented a second step of the influence process. The second step assures that the number of arguments that are relevant for an agent remains constant at S during the whole sense making process. This implies that when an agent i^* has adopted an argument that has not been relevant before, one of the arguments that are currently relevant for i^* will be dropped. We assume that agents drop the argument that has been adopted least recently. This reflects the idea that every time an agent hears an argument from an interaction partner, the cognitive importance of that argument is reinforced. The longer ago an argument has been heard from another agent for the last time, the less important the argument is considered to be and sooner or later it will be seen as entirely unimportant. Technically, we implement this in the model such that the relevance matrix of i^* is updated by subtracting one from all non-zero recency values. The argument that was communicated between i^* and j^* in the present event adopts at the end of the iteration a recency value of S ($s_{k^*,l^*,i^*} = S$). All other relevant arguments decline in recency. We also have tested alternative dropping rules, to assure the robustness of our results. Most importantly, we implemented that the argument for dropping is selected at random. Computational experiments revealed that all qualitative results reported below are robust to this modification of the model.

To illustrate the updating phase, Figure IV.2 contains two examples. Assume that matrix (b) is the initial relevancy matrix of agent *i**. Matrix (c) is the relevancy matrix of *i**'s interaction partner *j**. Before the update, the first argument is not relevant for *i** (see the circle in matrix (b)), but it is relevant for *j**. Hence, it is possible that *i** adopts the first argument, resulting in the updated relevance matrix for *i** shown in (d). Here the communicated argument is maximally recent (see the circle in matrix d). The recency of the remaining arguments has been reduced by 1. Note that this changed *i**'s opinion which shifted from -1/3 to +1/3 because *i** adopted a pro and dropped a con argument. As a second example, assume now that not the first argument is selected for update but argument number 4. This argument has already been relevant for *i** (see the square in matrix (b)). However, its recency has increased due to the interaction with *j** (see the square in matrix (e)). Note that this has no consequence on *i**'s opinion.

Interaction events are iterated until the system reaches equilibrium (Young 2001). Our model has exactly two equilibria, corresponding to perfect consensus or perfect subgroup polarization. Perfect consensus is reached when all agents hold the same opinions and, base these opinions on the same arguments. Then, no further change is possible. If all agents hold the same opinion but base that opinion on different arguments, then opinions can still change in upcoming events. Perfect subgroup polarization obtains if there are two subgroups, the members of each subgroup agree on all opinions and arguments with each other and the pairwise similarity ($sim_{i,j}$) between agents of different subgroups is zero. That is, the members of the subgroups maximally differ with respect to all demographic attributes and all opinions. If all members of the subgroups base their opinions on the same arguments, then this outcome is stable.

Obviously, the second equilibrium can only be reached in teams where faultline strength is maximal (f=1), because in these teams there are no crisscrossing agents. Crisscrossing agents share at least one demographic attribute with members of both demographic subgroups. Accordingly, if there is a crisscrossing agent and the two subgroups still disagree, a positive probability remains that arguments of the one subgroup are adopted by the other and the disagreement will vanish.

Some of our propositions and experiments focus on the duration of the sense making process, i.e. the time that it takes before consensus or perfect polarization has been reached. To be sure, we refrain from formulating statements about effects of the independent variables in our experiments on the *absolute* duration (e.g. in days or seconds) of the sense making process. We are not aware of any empirical evidence that would allow assessing meaningfully the duration of a simulated interaction event in real time. However, it seems reasonable to assume that the more interaction events occur before equilibrium, the longer such a process also would take in real time. This allows us to compare the length of the process in terms of number of events under different conditions.

IV.4. Simulation Experiments

The central outcome variable of faultline theory is the level of subgroup polarization in work teams. To quantify subgroup polarization we use a measure called *polarization* (Flache and Mäs 2008a). It captures the degree to which the group can be separated into a small set of factions who are mutually antagonistic in the opinion space and have maximal internal agreement. To compute *polarization*, we use the variance of pairwise opinion agreement across all pairs of agents in the population, where agreement is ranging between -1 (total disagreement) and +1 (full agreement), measured as one minus the average distance of opinions (averaged across all K subdimensions). This measure obviously adopts its lowest level of zero for the case of perfect opinion consensus. The maximum level of opinion polarization (*polarization*=1) is obtained when the population is equally divided between the opposite ends of the opinion scale at -1 and +1 and all opinion dimensions are perfectly correlated.

To test whether our propositions follow consistently from the model, we conducted computational experiments varying three model parameters: the strength of faultlines (f), the initial correlation between demographic attributes and opinions (w) and the strength of homophily (h). For the remaining parameters we imposed values that are realistic and allow at the same time that the model generates sufficient variance in the outcome variables. Across all conditions, we assumed a team size of 20 (N=20) and used three fixed (demographic) attributes (D=3). Furthermore we assume that only one issue is relevant (K=1). Including further issues (K>1) makes it necessary to control for the correlations between the opinions and between each opinion and the demographic attributes. Since we focus here on the effects of demographic faultlines, we decided to keep the number of parameters varied in the experiment low and consider only one issue. Finally, we assumed that there exist always 10 pro (P=10) and 10 con (C=10) arguments. We assigned the value 4 to S in all conditions meaning that the actors base their opinions on 4 arguments.

To vary faultline strength (f) in the experiments, we used exactly the same distributions of demographic attributes that we used in Figure IV.1 (cf. Flache and Mäs

2008a; 2008b). We varied the Pearson correlation between each pair of demographic attributes from 0 to 1 in steps of .2. Of course, there are many alternative distributions of the three variables that result in the same *bivariate* correlations. The distributions we used, however, are the only ones that produce equal correlations between all pairs of demographic attributes and at the same time keep diversity maximal. We chose equal correlations to resolve a conceptual unclarity in Lau and Murnighan's definition of faultline strength. Do we, for example, speak of a strong faultline if two variables x and y are perfectly correlated but completely unrelated to a third variable z? Or, would we regard the faultline as stronger or weaker if x and y are correlated only with r=.8 but the correlation between x and z would rise to .6? These questions do not occur if all pairs or variables are equally correlated. Furthermore, considering unequal correlations between the demographic attributes does not affect the results. Also with unequal correlations, it holds that the weaker the correlation between the demographic dimensions, the smaller are the subgroups and the more crisscrossing actors there are in a team. Furthermore, as long as not all pairwise correlations are maximal (f=1), crisscrossing actors must be present.

To manipulate the level of initial congruency (w), we related the initial opinion to the first demographic attribute. The extent to which this affects the correlation of the opinion with the remaining demographic attributes depends on faultline strength (f). The stronger the faultline, the higher is the correlation between the first demographic attribute and the other demographic attributes. Accordingly, the stronger the faultline, the more similar are the correlations between the opinion and the first, second etc. demographic attribute. Technically, we assigned S arguments to each agent. For each of the S arguments, we assigned one of the existing pro arguments with the probability w when the agent holds the value 1 at the first demographic attribute and one of the con arguments otherwise. Agents with the value -1 at the first demographic attribute received a pro argument with probability 1-w. For instance, if w is 0.5, then pro and con arguments always have the same probability to be assigned. On average, this results in a Pearson-correlation between the first demographic attribute and the opinion of zero. However, as w increases, agents with the value 1 (-1) at the first demographic attribute more likely receive a pro (con) argument. This entails a higher Pearson-correlation between the first demographic attribute and the opinion as w increases. Under w=1, the opinion and the first demographic attribute perfectly align. More precisely, all agents that hold the value 1 at the first demographic attribute also hold opinion values of 1 and all agents who belong to the other demographic subgroup on the first dimension, hold opinion values of -1.

We varied *w* between .5 and 1 in steps of .1. We do not consider *w*-values below .5. Such values would lead to a negative correlation between the opinion and the demographic attributes. Since the actual values of the opinion and the demographic attributes have no substantial meaning, it makes no difference if opinions and demographic are positively or negatively correlated. To test the effects of the strength of homophily we manipulated the parameter *h* (see equation 3), varying it between 1 and 5 steps of 1. A value of *h*=1 expresses that agents have a weak preference to interact with similar team mates. The value of *h*=5 corresponds to a very strong homophily.

All in all we inspect 6*6*5 = 180 conditions in our computational experiments. The 5 conditions in which faultline strength is maximal (*f*=1) and the initial correlation between opinions and demographic attributes is maximal (*w*=1) have been excluded because the similarity ($sim_{i,j}$) between agents is under this condition either 1 or 0. In these cases, it is logically impossible that members of different subgroups will interact and opinions will therefore not change. For reliability, we conducted 500 independent replications per condition.

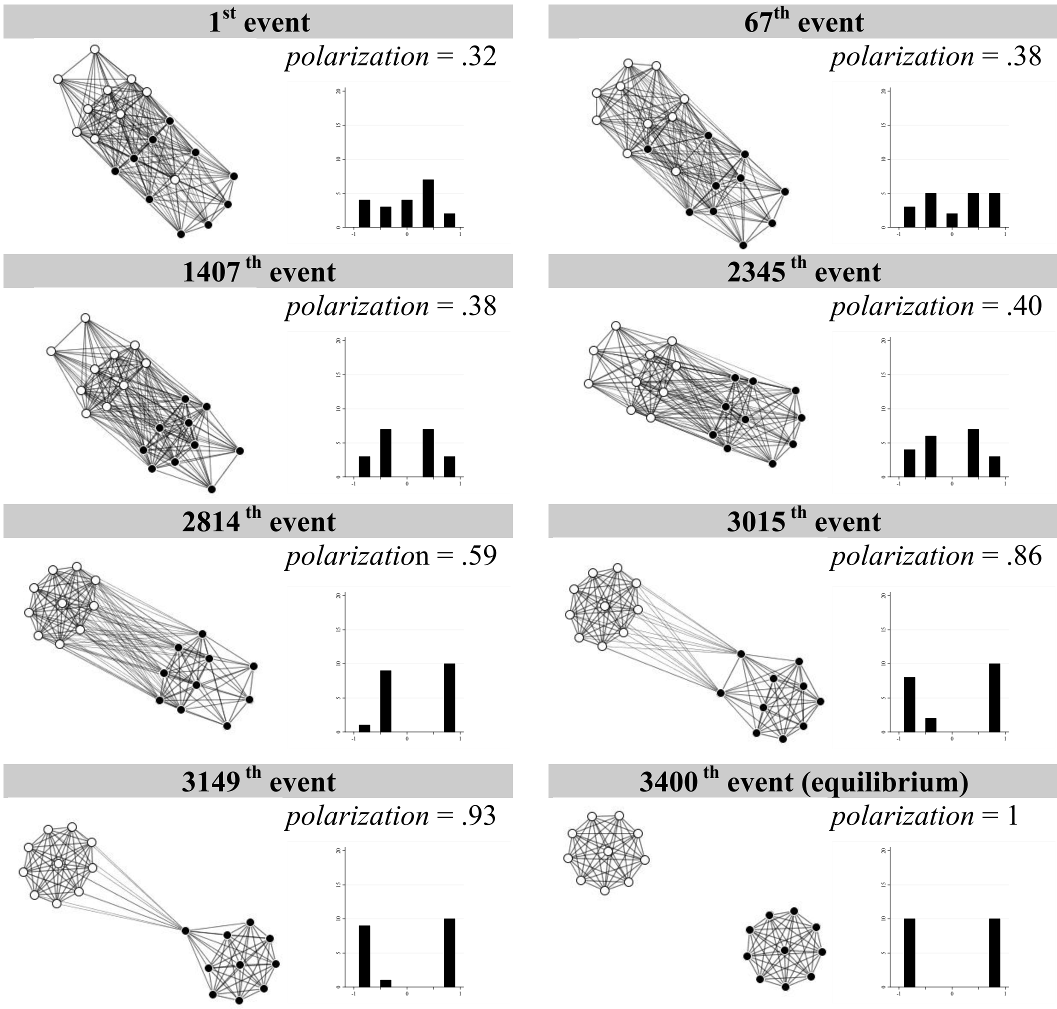
IV.5. Results

We present the results in three steps. In the first step, we present two ideal-typical simulation runs to illustrate model dynamics. We then turn to the consistency tests of the four propositions. Finally, we present additional analyses that revealed an unexpected effect of faultline strength.

IV.5.1. Ideal-typical simulation runs

Figure IV.3 demonstrates an ideal-typical simulation run with maximal faultline strength (*f*=1). To trigger subgroup polarization, we imposed conditions that, according to the propositions, make polarization very likely. We assumed strong homophily (*h*=5) and imposed a relatively strong correlation of initial opinions with demographic attributes (*w*=.8). The latter generated for this run an initial Pearson correlation between the opinion and the three demographic attributes of .77. Figure IV.3 shows the development of *polarization* and the distribution of the opinion at different stages of the simulation run. The histograms show the respective opinion distribution. The network pictures describe the resulting interaction structure. In the network pictures, each agent is represented by a circle. The color of a circle indicates to which of the two demographic subgroups the respective agent belongs. Each pair of agents that has a nonzero overall similarity ($sim_{i,j}$) is connected by a line, symbolizing that there is a nonzero probability that these two agents interact. Because we focus here on the development of opinions in the team, the arrangement of the circles is only based on opinion similarity. Thus, circles are arranged in a way such that the nearer agents are placed to each other, the more similar their opinions are (Kamada and Kawai 1989; McFarland and Bender-deMoll 2007).

Figure IV.3: Ideal-typical run with maximal faultline strength (*f*=1, *h*=5, *w*=.8)



Initially (1st event in Figure IV.3), the opinion was almost uniformly distributed in this simulation run. Nevertheless, the corresponding network picture reveals that there are already initially systematic opinion differences between the demographic subgroups. The change of the histograms of the subsequent events shows that over time opinion differences between the subgroups increase. Consequently, the number of lines between the subgroups also decreases over time. Eventually (by event 3400) the subgroups hold maximally opposing opinions. The exchange of arguments between subgroups stopped at this point, because there is neither an overlap in demographic attributes nor in opinions between agents from different subgroups. Opinion changes have now become impossible because agents only interact with team members hold the same opinion and arguments.

Figure IV.4: Ideal-typical run with 3 crisscrossing agents (*f*=.8, *h*=5, *w*=.8)

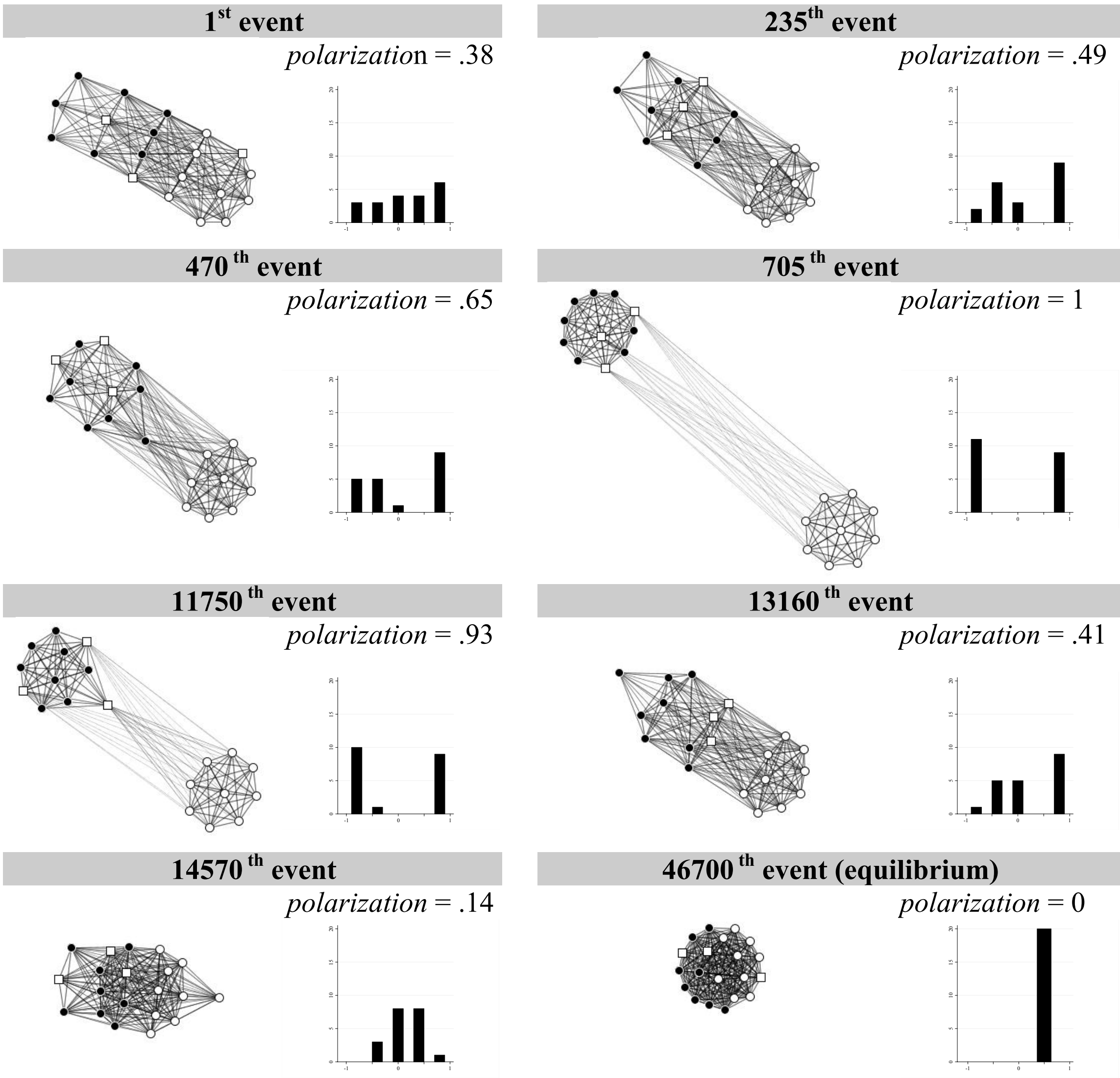


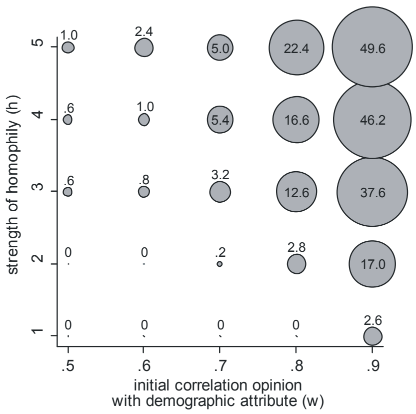
Figure IV.3 depicts ideal-typical dynamics that ended in a stable group split. This shows that our model can generate the dynamics that Lau and Murnighan described in their informal reasoning if the faultline is maximally strong. Proposition 2, however, expects that dynamics differ crucially when crisscrossing actors are present. Figure IV.4 shows an ideal-typical run that supports the proposition. In this run, faultlines were slightly weaker than in the condition of Figure IV.3. For comparison, we retain all further parameters of the first illustrative run with maximal faultline strength (*h*=5, *w*=0.8), but we slightly reduce the strength of the faultline to f=0.8. Now, the team contains three crisscrossing agents (see the three squares in network pictures of Figure IV.4). Initially (see 1st event), the opinion is again almost uniformly distributed and the demographic subgroups already hold somewhat different opinions. Again, we observe increasing subgroup polarization, just as Lau and Murnighan proposed. After the 705th event the work teams fell apart into two opposing subgroups with maximally different opinions. Within the two subgroups, the agents share the same opinions and also quickly coordinate on a common vector of arguments. However, the two subgroups are not completely unconnected: due to the three crisscrossing actors there is still some exchange of arguments between the subgroups. The network picture of the 11750th event demonstrates that one of the crisscrossing actors adopted an argument that changed his opinion. Subsequently, this argument spreads in the crisscrossing actors' subgroup and the opinion differences between the subgroups decrease (see event 13160). This convergence process continues until overall consensus is reached.

IV.5.2. Consistency tests of the propositions

Long- term effects. According to Lau and Murnighan, higher faultline strength entails more subgroup polarization (proposition 1). Proposition 2, however, claims that in teams with non-maximal faultline strength this effect can only be observed in the short term. Our experiments clearly confirmed proposition 2. All simulated work teams with faultline strength below its theoretical maximum eventually ended in overall opinion consensus. That is, all team members held the same opinion and based it on exactly the same arguments.

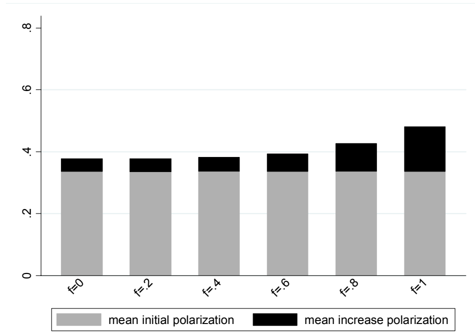
In teams with maximally strong faultlines, however, we found perfect subgroup polarization, but not in all simulation runs. Figure IV.5 shows how the initial congruency (w) and the strength of homophily (h) affected the frequency of runs that ended in perfect subgroup polarization. More precisely, the size of the bubbles in Figure IV.5 corresponds to the percentage of runs under the respective condition that ended in stable group splits. For instance, 49.6% of 500 runs with very strong homophily (h=5) and very strong initial correlation between the opinion and the demographic attributes (w=.9) ended in a group split with two equally large subgroups and maximally opposing opinions. As suggested by the propositions 3 and 4, the higher the values of h and w are, the more likely subgroup polarization occurs. Figure IV.5 thus confirms that our implementation of Lau and Murnighan's mechanisms can explain stable subgroup polarization in teams with maximally strong faultlines. But, even with maximal faultline strength, group splits remain unlikely if homophily is weak or the opinions are not already initially strongly aligned with the demographic attributes.

Figure IV.5: Percentage of runs that ended in stable splits (*f*=1)



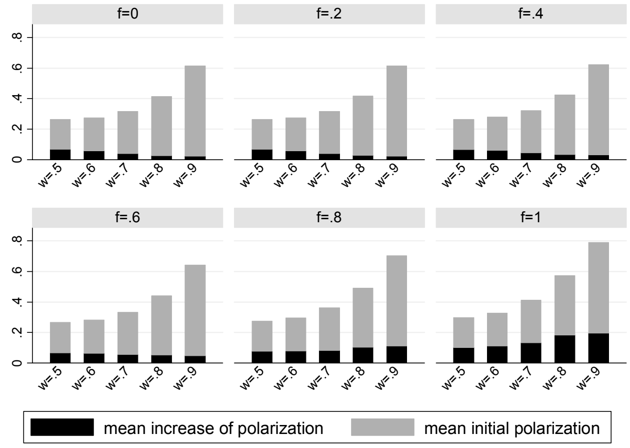
Short-term effects. Proposition 2 predicts increasing subgroup polarization in the short term even though teams reach consensus in the long term. To asses the short term polarization in the simulated work teams we measured the maximal value of *polarization* that teams exhibited during simulation runs and compared this value to the runs' initial value of polarization. The difference between these two values indicates to which degree groups split up in the short run independent on whether the split occurred right at the beginning of the run or later. To test if faultlines trigger short term polarization (proposition 2), Figure IV.6 shows bar graphs broken down by faultline strength (*f*). The gray part of each bar in Figure IV.6 depicts the average initial level of *polarization* in the teams. The black part of the bars shows the average increase in *polarization*. Both parts add up to the average of the maximal value of *polarization*. We excluded the conditions where *w*=1, because here *polarization* is initialized at it's logical maximum and cannot further increase. Figure IV.6 shows a stronger increase in the maximal value of *polarization* as faultlines become stronger (*f*) and thus supports proposition 2. At least in the short run, faultlines trigger subgroup polarization.

Figure IV.6: Average maximal opinion *polarization* over *f*, (15000 runs per bar)



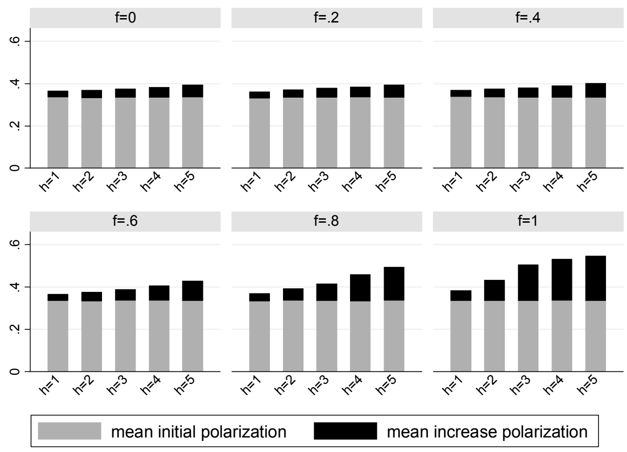
According to proposition 3, the higher the initial congruency, the stronger should be short term effects of strong faultlines on polarization. To test that, we display in Figure IV.7 the effects of congruency (*w*) on maximal *polarization* broken down by faultline strength *f*. The figure shows that the initial *polarization* depends on *w* (see the gray parts of the bars). This is a technical consequence of congruency that occurs because opinions align closer with the 50:50 split on the values of +1 and -1 in the first demographic attribute, as *w* increases. There is, however, no such relationship of initial opinion polarization to faultline strength because the distribution of the first demographic attribute is the same for all levels of faultline strength. It turns out that the maximal value of *polarization* (see the complete bars) in all subgraphs increases with *w*. However, as the size of the black areas shows, this is mainly the result our manipulation of *w*. If faultlines are not strong (f<.8), the mean increase of *polarization* declines with the initial correlation between opinion and the first demographic attribute. We believe that this results from a ceiling effect. If faultlines are weak, then most pairs of agents have a relatively high similarity (sim_{ij}) because of shared demographic attributes. The potential of opinion polarization in these teams is thus very low. If *w* is high, these teams start out close to their potential maximum of *polarization*. As a consequence, *polarization* can only rise moderately above the initial level and will decline soon thereafter. If faultlines are strong (f>.6), however, the model produces the effect of *w* that Proposition 3 expected. The black parts of the bars in the subgraphs for faultline strengths of .8 and 1 show that a higher initial correlation of opinion and first demographic attribute entails more opinion polarization.

Figure IV.7: Average maximal opinion *polarization* over *w*, by *f* (2500 runs per bar)



Proposition 4 suggests that subgroup polarization should increase with stronger homophily. Figure IV.8 confirms that the increase in *polarization* (see the black parts of the bars) is higher for stronger homophily (*h*). Comparison of different faultline levels also reveals that the effect of homophily strength increases in the strength of faultlines (*f*). If faultlines are weak, then even a very strong preference of the agents to interact with similar team members will cause only little increase in *polarization* in the short run. If faultlines are stronger, then strong homophily results in a larger increase in *polarization*.

Figure IV.8: Average maximal opinion *polarization* over *h*, by *f* (2500 runs per bar)

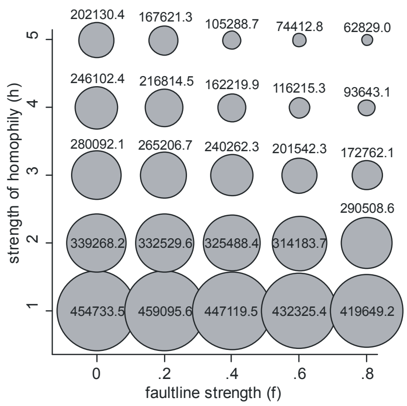


IV.5.3. Relative time until convergence

The simulation experiments have confirmed that all teams that contain crisscrossing actors eventually arrived at consensus, even though many polarized in the short term. The analyses of the length of this convergence process, however, led to an unexpected and counter-intuitive result: the stronger the faultline in a team and the stronger homophily, the *faster* the teams arrive at consensus. Figure IV.9 shows a bubble graph expressing the average number of events it took until the runs ended in overall consensus, broken down by faultline strength (*f*) and homophily strength (*h*). The graph shows that the less events were needed to reach consensus, the stronger the faultline was. It also shows that stronger homophily is associated with faster emergence of opinion consensus.

To confirm this counter-intuitive result, we conducted simulation experiments where teams started with perfect polarization (w=1) and varied faultline strength. In the runs with weak faultlines (f=0) the teams very quickly overcame the group split but it took them very long to arrive at consensus. By contrast, it took the teams with a strong faultline (f=.8) longer to overcome the initial group split. However, once the split was overcome, the teams quickly found a consensus.

Figure IV.9: Average number of events until the teams arrived at an overall consensus



We explain this effect as a consequence of the interaction structure in teams with strong faultlines. The same interaction structure that causes subgroup polarization in the short term accelerates the convergence process as soon as the opinion split has been overcome. As we have shown in Figure IV.1, teams with strong faultlines consist of subgroups. Frequent exchange of arguments within the subgroups makes the subgroups reach internal consensus quickly. When there are crisscrossing agents, however, from time to time new arguments enter a subgroup and lead to changing opinions and a new subgroup consensus. If there are only a few crisscrossing agents present (strong faultlines), this process leads to a gradual convergence of opinions across the subgroups. Most importantly, this accelerates coordination on a single vector of arguments in the whole team much faster. The reason is that subgroups adopt a new argument and drop one of the arguments used before. Once an argument is dropped by a subgroup, it will not reoccur in later interactions. With weak faultlines, however, there are more crisscrossing agents and new arguments enter the discussion within subgroups more frequently. This can be so frequent that subgroups do not manage to find consensus before a new argument enters. As a consequence, the number of arguments and opinion diversity within each subgroup remains high and the gradual convergence of subgroups that we found in groups with strong faultlines does not develop. Furthermore, frequent argument exchange with crisscrossing actors leads to a fast spread of arguments across the entire team. Thus, if a subgroup collectively drops an argument, this argument may still be used by other team members and might re-enter the discussion in the subgroup over and over again. Overall, the convergence of opinions occurs faster in the structured interaction network of a team with a strong faultline than in the unstructured communication pattern in a team with a weak faultline.

IV.6. Summary and Implications

Lau and Murnighan (1998) argued that teams with a strong demographic faultline likely experience subgroup polarization. We challenged this prediction, arguing that Lau and Murnighan overlooked the important role of crisscrossing actors in the sense making process of teams. Crisscrossing actors are team members who share some demographic attributes with multiple subgroups and can thus function as a bridge over the faultline. We showed that the faultline concept implies that even teams with very strong faultlines comprise at least a few crisscrossing actors. Accordingly, we argued that also in teams with strong faultlines there are processes that could prevent subgroup polarization or, if teams are polarized, help to overcome group splits. This led us to propose that strong faultlines breed subgroup polarization only in the short run. If there are crisscrossing actors in a team, even teams with strong faultlines will eventually overcome polarization. Moreover, we propose that Lau and Murnighan's theory implicitly factors crisscrossing effects in, although they did consider this explicitly. To underpin this claim, we developed a formal model based on the central behavioral assumptions of Lau and Murnighan's theory. We conducted computational experiments to test whether our new propositions follow consistently from the behavioral assumptions. Our analyses clearly confirmed this. We also proposed that faultline effects may crucially depend on core assumptions hidden in previous theoretical elaborations. We found that stronger faultlines only imply opinion polarization if demographic attributes are strongly correlated with the opinions of team members even before they influence each other. Moreover, to logically derive effects of strong faultlines, the assumption is needed that homophilious selection plays an important role in interactions within the team. Finally, contrary to intuition, our simulations revealed that teams with strong faultlines might be faster in arriving at an opinion consensus.

Our analyses confirm that teams with strong faultlines experience more polarization than teams with weak faultlines. However, if faultlines are not maximally strong, effects of faultline strength occur only for the short term dynamics in a team. In the long run, group splits disappear sooner or later. This appears to be good news for managers. Nevertheless, we advise readers to interpret our results with caution. The main purpose of our analysis was to point to hidden implications of the mechanisms assumed by faultline theory. This does not preclude that other mechanisms not considered by the theory may lead to different consequences. Specifically, our formal model did not consider the possibility that social identities form around subgroups in the process of a group split. Members of the subgroups may then "act to legitimize the subgroups, and conflict between them may continue to be likely" (Lau and Murnighan 1998: 333). Strong subgroup identification may motivate team members to refuse communication with crisscrossing actors. Identification might also promote the development of stereotypes about the demographic subgroups. Since crisscrossing actors fit into none of the stereotypes, they may be rejected by members of both demographic subgroups. If such negativity arises then crisscrossing actors will not be able to conciliate.

Despite the possibility that identity formation may reduce the influence of crisscrossing actors, our results should also not be discarded too readily. We have shown that integrating effects of crisscrossing actors can in the long run only be precluded if these actors are perfectly excluded from the interaction networks within the subgroups. Even if subgroup identities form, it seems a rather extreme assumption that they can entirely prevent any subgroup influence via crisscrossing actors. It seems more plausible that the strength of subgroup identities affects how long it takes until the initial group splits can be overcome, but not the eventual outcome given the team has enough time to converge to a consensus. This also suggests that in the actual practice of work teams, crisscrossing actors may be important to overcome the negative effects of faultlines if management succeeds in creating conditions that support their integrating role. For example, an amicable and friendly environment in the work team may be important to reduce subgroup identifications and may therefore facilitate the exchange of arguments between the subgroups via crisscrossing actors.

This paper revealed that short term consequences of group dynamics might crucially differ from their effects in the long run. Other recent contributions also proposed effects of time on consequences of demographic diversity in work groups. Most prominently, Harrison et al. (2002) argued that as team members get to know each other, the relevance of surface level (demographic) characteristics will diminish and members will base selection of interaction partners more on psychological similarity (personality, values, attitudes, beliefs). Like our reasoning, their argument suggests that the impact of demographic diversity and thus of demographic faultlines declines over time (see also: Pelled, Eisenhardt and Xin 1999). However, we have shown that this follows already from the elementary behavioral assumptions of faultline theory, without the need to necessarily include additional mechanisms such as the distinction between surface similarity and psychological similarity. This demonstrates that already relatively simple models of social processes can be too complex to grasp their logical consequences by informal reasoning. Formal methods, therefore, are useful to study such complex systems and to reveal unexpected consequences of theories that may remain undiscovered otherwise.

V. Individualization and Clustering[15]

Abstract
The models which we have discussed so far can explain opinion polarization. However, they fail to explain clustering. In this chapter, we will focus on the clustering problem. Existing social influence models can generate clustering when the assumption is included that individuals refuse to interact with others that are too dissimilar. However, we will show that clustering breaks down when this assumption is slightly relaxed. A possible solution to this problem, based on classical sociology, incorporates striving for individualization. However, we will show that within the existing modeling framework, this also fails to explain opinion clustering. In particular, we will demonstrate that allowing tiny individual opinion perturbations (white noise) triggers social influence cascades that inevitably lead to monoculture. However, increasing noise implies rampant individualism rather than clustering.
We will present a new solution to the clustering problem and show that the new model does generate clustering. The key element of our model is the assumption that individuals will increase their uniqueness only when many others hold similar opinions. Individuals who do not share an opinion with many others are already unique and do not need to individualize.
In a computational experiment, we will identify conditions under which populations split up into homogenous clusters with distinct opinions. Once they have developed, clusters are temporarily stable but may merge. Merging, however, triggers the development of new clusters. In this way, clustering is a stable outcome. Paradoxically, the new model's predictions are not only robust to noise, but noise is the central mechanism that triggers cluster formation.

V.1. Introduction

Modern societies are characterized by a large degree of pluralism in social, political and cultural opinions (Fiorina and Abrams 2008). Yet, individuals typically interact within small communities. Recently, the clustering of social interactions (Barabasi and Albert 1999; Palla, Barabasi and Vicsek 2007; Watts and Strogatz 1998) has been extensively studied, from email communication (Liben-Nowell and Kleinberg 2008) to phone calls

[15] This chapter was written together with Andreas Flache and Dirk Helbing. An elaborated version of this chapter has been accepted for publication at PLoS Computational Biology. The title of the published article is: "Individualization as driving force of clustering phenomena in humans". It will be freely available online (www.ploscompbiol.org).

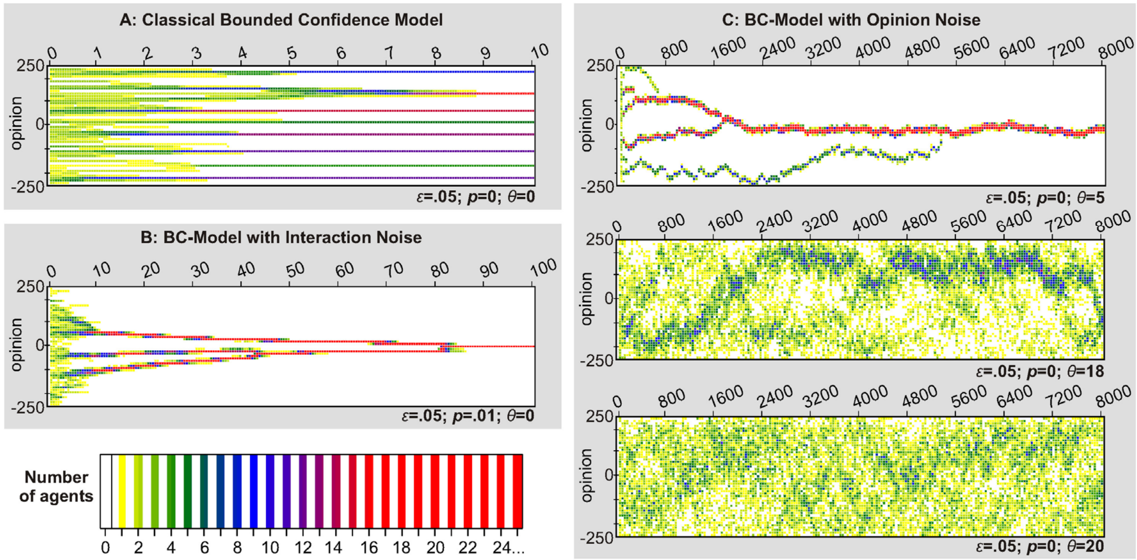
(Palla, Barabasi and Vicsek 2007) to scientific collaboration networks (Newman 2004), or sexual contacts (Liljeros et al. 2001). It is much less understood, however, how the pattern of opinion clustering that characterizes pluralism comes about.

Empirical studies suggest that opinions differ globally, while they cluster locally within geographical regions (Glaeser and Ward 2006), socio-demographic groups (Mark 2003), or internet communities (Lazer et al. 2009). The lack of a theoretical understanding of opinion clustering is pressing, since both, local consensus and global diversity are precarious. On the one hand, cultural diversity may get lost in a world where people are increasingly exposed to influences from mass media, Internet communication, interregional migration, and mass tourism, which may promote a universal monoculture (Friedman 2005; Greig 2002), as the extinction of languages suggests (Sutherland 2003). On the other hand, increasing individualization threatens to disintegrate the social structures in which individuals are embedded, with the possible consequence of the loss of societal consensus (Beck 1994; Durkheim 1997 [1893]). This is illustrated by the decline of the social capital binding individuals into local communities (McPherson, Smith-Lovin and Brashears 2006).

Early formal models of social influence imply that monoculture is unavoidable, unless a subset of the population is perfectly cut off from outside influences (Abelson 1964). Social isolation, however, appears questionable as explanation of pluralism. In modern societies, distances in social networks are quite short on the whole, and only relatively few random links are required to dramatically reduce network distance (Watts and Strogatz 1998).

Aiming to explain pluralism, researchers have incorporated the empirically well-supported observation of “homophily”, i.e. the tendency of “birds of a feather to flock together” (Aral, Muchnik and Sundararajan 2009; McPherson, Smith-Lovin and Cook 2001), into formal models of social influence (Nowak, Szamrej and Latané 1990). These models typically assume ``bounded confidence'' (BC) in the sense that only those individuals interact, whose opinions do not differ more than a given threshold level (Deffuant, Huet and Amblard 2005; Hegselmann, Flache and Möller 2002). As Figure V.1A illustrates, BC generates opinion clustering, a result that generalizes to model variants with categorical rather than continuous opinions (Axelrod 1997; Nowak, Szamrej and Latané 1990). However, clustering in the BC-model is sensitive to “interaction noise”: A small random chance that agents may interact even when their opinions are not similar, causes monoculture again (see Figure V.V. 1B).

Figure V.1: Typical opinion dynamics generated by the Bounded confidence model



Opinion dynamics produced by the bounded confidence (BC) model (Hegselmann and Krause 2002) with and without noise for populations of 100 agents on an opinion scale ranging from -250 to 250. Initial opinions are uniformly distributed. For visualization, the opinion scale is divided into 50 bins of equal size. Color coding indicates the relative frequency of agents in each bin. **(A)** Dynamics of the BC-model without noise (Hegselmann and Krause 2002) over 10 iterations. At each simulation event, one agent's opinion is replaced by the average opinion of those other agents who hold opinions o_j within the focal agent's confidence interval (o_i-$\varepsilon < o_j < o_i+\varepsilon$). For ε=0.05, one finds several homogeneous clusters, which stabilize when the distance between all clusters exceeds the confidence threshold ε. **(B)** Computer simulation of the same BC-model, but considering interaction noise. Agents that would otherwise not have been influential, now influence the focal agent's opinion with a probability of p=0.01. This small noise is sufficient to eventually generate monoculture. **(C)** Simulation of the BC-model with opinion noise. After each opinion update, a normally distributed random value drawn from $N(0,\theta)$ is added to the opinion. Under weak opinion noise (θ=5), one cluster is formed, which carries out a random walk on the opinion scale. When the opinion noise is significantly increased (θ=18), there is still one big cluster, but many separated agents exist as well (cf. Figure V.4). With even stronger opinion noise (θ=20), the opinion distribution becomes completely random.

To avoid this convergence of opinions, it was suggested that individuals would separate themselves from negatively evaluated others (Macy et al. 2003; Mark 2003). However, recent empirical results do not support such "negative influence" (Krizan and Baron 2007). Scientists also tried to avoid convergence by "opinion noise", i.e. random influences, which lead to arbitrary opinion changes with a small probability. Similar to earlier modeling work on *nominal* social-influence models (Carley 1991; Klemm et al. 2003a; Klemm et al. 2003b; Mark 1998; Mark 2003), Pineda et al. (2009) included uniformly distributed opinions noise. This implementation of opinion noise assumes sudden, large, and unmotivated opinion changes of individuals. However, theories of social integration (Beck 1994; Durkheim 1997 [1893]; Hornsey and Jetten 2004; Vignoles, Chryssochoou and Breakwell 2000) and empirical studies of individualization (Imhoff and Erb 2009; Snyder and Fromkin 1980) show a tendency of incremental opinion changes rather than arbitrary opinion jumps. Incremental opinion changes, however, tend to promote monoculture, even in models with categorical rather than continuous opinions (Klemm et al. 2003a). Figure V.1 demonstrates that adding a “white noise” term ($N(0,\theta)$) to an agent's current opinion in the BC model fails to explain opinion clustering. Weak opinion noise (θ=5) triggers convergence cascades that inevitably end in monoculture. Stronger noise restores opinion diversity, but not pluralism. Instead, diversity is based on frequent individual deviations from a predominant opinion cluster (for θ=18). However, additional clusters can not form and persist, because opinion noise needs to be strong to separate enough agents from the majority cluster - so strong that randomly emerging smaller clusters cannot stabilize.

In conclusion, the formation of persistent opinion clusters is such a difficult puzzle that all attempts to explain them had to make assumptions that are difficult to justify by empirical evidence. The solution proposed in the following, in contrast, builds on sociological and psychological research. The key innovation is to integrate another decisive feature into the model, namely the “striving for uniqueness” (Imhoff and Erb 2009; Snyder and Fromkin 1980). While individuals are influenced by their social environment, they also show a desire to increase the uniqueness when too many other members of society hold similar opinions. This suggests that noise in individual opinion formation is adaptive, creating a dynamic interplay of the integrating and disintegrating forces highlighted by Durkheim's classic theory of social integration (Durkheim 1997 [1893]). Durkheim argued that integrating forces bind individuals to society, motivating them to conform and adopt values and norms that are similar to those of others. But he also saw societal integration as being threatened by disintegrating forces that foster individualization and drive actors to differentiate from one another (Beck 1994; Hornsey and Jetten 2004; Vignoles, Chryssochoou and Breakwell 2000). The continuous “Durkheimian opinion dynamics model”' proposed in the following can explain pluralism, although it incorporates all the features that have previously been found to *undermine* clustering: (1) a fully connected influence network, (2) absence of bounded confidence, (3) no negative influence, and (4) white opinion noise. From a methodological viewpoint, our model builds on concepts from statistical physics, namely the phenomenon of ”nucleation”(Stanley 1971), illustrated by the formation of water droplets in supersaturated vapor. However, by assuming adaptive noise, we move beyond conventional nucleation models.

Computational experiments reveal that our model generates pluralism as an intermediate phase between monoculture and individualism. When the integrating forces are too strong, the model dynamics inevitably implies monoculture, even when the individual opinions are initially distributed at random. When the disintegrating forces prevail, the result is what Durkheim called "anomie", a state of extreme individualism without a social structure, even if there is perfect consensus in the beginning. Interestingly, there is no sharp transition between these two phases, when the relative strength of both forces is changed. Instead, we observe an additional, intermediate regime, where opinion clustering occurs, which is independent of the initial condition. In this regime, adaptive noise entails robust pluralism that is stabilized by the adaptiveness of cluster size. When clusters are small, individualization tendencies are too weak to prohibit a fusion of clusters. However, when clusters grow large, individualization increases in strength, which triggers a splitting into smaller clusters ("fission"). In this way, our model solves the cluster formation problem of earlier models. While in BC models, white noise causes either monoculture or fragmentation (Figure V.1C), in the Durkheimian opinion dynamics model proposed here, it *enables* clustering. Therefore, rather than *endangering* cluster formation, noise supports it. In the following, we describe the model and identify conditions under which pluralism can flourish.

V.2. The Model

The model has been elaborated as an agent-based model (Bonabeau 2002) addressing the opinion dynamics of interacting individuals. The simulated population consists of N agents i, representing individuals, each characterized by an opinion $o_i(t)$ at time t. The numerical value for the opinion varies between a given minimum and maximum value on a metric scale.

We use the term "opinion" here, for consistency with the literature on social influence models. However, o_i may also reflect behaviors, beliefs, norms, customs or any other cardinal cultural attribute that individuals consider relevant and that is changed by social influence. The dynamics is modeled as a sequence of events. Every time $t'=k/N$, the computer randomly picks an agent i and changes the opinion o_i by the amount

$$\Delta o_i = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \left(o_j - o_i\right) w_{ij}}{\sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij}} + \xi_i \tag{1}$$

The first term on the rhs of Eq. 1 models the integrating forces of Durkheim's theory. Technically, agents tend to adopt the weighted average of the opinions o_j of all other members j of the population. Implementing homophily, the social influence w_{ij} that agent j has on agent i is the stronger, the smaller their opinion distance $d_{ij} = |o_j - o_i|$ is. Formally, we assume

$$w_{ij} = e^{-d_{ij}/A} = e^{-|o_j - o_i|/A} \tag{2}$$

The parameter A represents the range of social influence of agents. For small positive values of A, agents are very confident in their current opinion and are mainly influenced by individuals who hold very similar opinions, while markedly distinct opinions have little impact. The higher A is, however, the more are agents influenced by individuals with considerably different opinions and the stronger are the integrating forces in our Durkheimian theory.

$$\theta_{it} = s\sum_{j=1}^{N} e^{-d_{ij}} \tag{3}$$

The larger the standard deviation, the stronger are the individualization tendencies of an agent. Following Durkheim's theory, equation 3 implements the assumption that an agent's striving for individualization is weak, if there are only a few others with similar opinions. Under such conditions, there is no need to increase distinctiveness. However, if many others hold a similar opinion, then individuals are more motivated to differ from others.

By including the focal agent i in the sum of Eq. 3, we assume that there is always some opinion noise, even when agent i holds a perfectly unique opinion. These fluctuations may have a variety of reasons, such as misjudgements, trial-and-error behavior, or the influence of exogenous factors on the individual opinion.

We use the parameter s to vary the strength of the disintegrating forces in society. The higher the value of s, the higher is the standard deviation of the distribution, from which ξ_i is drawn, and the stronger are the disintegrating forces. Finally, to keep the opinions of the agents within the bounds of the opinion scale, we set the value of ξ_i to zero, if the bounds of the opinion space would be left otherwise.

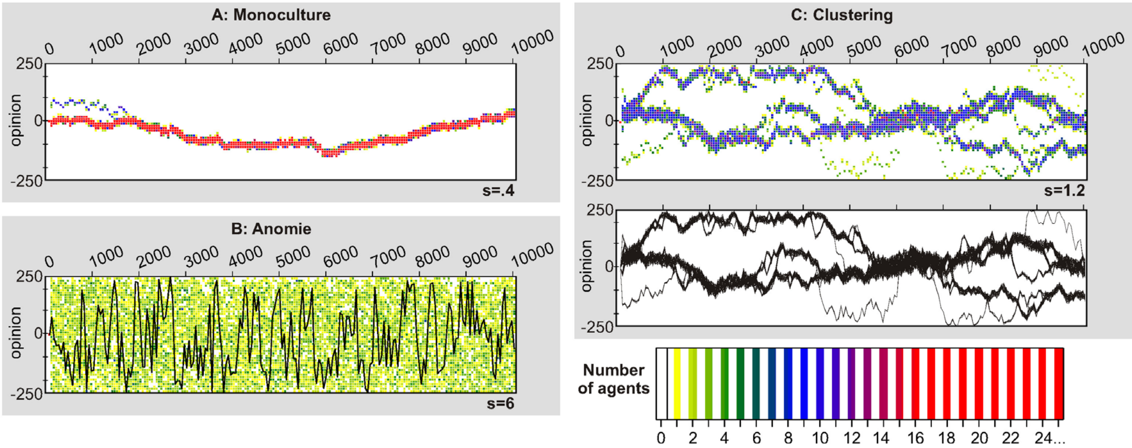
V.3. Results

We have studied the Durkheimian opinion dynamics model with extensive computer simulations, focussing ourselves on relatively small populations (N=100), because in this case it is reasonable to assume that all members may interact with each other. For bigger groups one would have to take into account the topology of the social interaction network as well. Such networks would most likely consist of segregated components ("communities"), which are not or only loosely connected with each other (Liben-Nowell and Kleinberg 2008; Liljeros et al. 2001; Newman 2004; Palla, Barabasi and Vicsek 2007). Because of the weak or missing connections *between* communities, it would not be so surprising if each community developed its own, shared opinion. In small, completely connected populations, however, the occurrence of diverse opinions is puzzling, as it cannot result from a lack of contacts between agents.

To illustrate the model dynamics, Figure V.2 shows three typical simulation runs for different strengths s of disintegrating forces, while the strength A=2 of the integrating force is kept constant. In each run, all agents start with an opinion in the middle of the opinion scale (o_i=0), i.e. conformity. This is an initial condition for which the classical BC-model does not produce diversity. Figure V.2A shows typical opinion trajectories for a population, in which the integrating forces are much stronger than the disintegrating forces. Consequently, the population develops collective consensus, i.e. the variation of opinions remains small, even though not all agents hold exactly the same opinion. Triggered by the random influences ξ_i, the average opinion performs a characteristic random walk.

When the disintegrating force prevails, the pattern is strikingly different. Figure V.2B shows that for large noise strengths s, the initial consensus breaks up quickly, and the agents' opinions are soon scattered across the entire opinion space.

Figure V.2: Typical opinion dynamics generated by the Durkheimian model

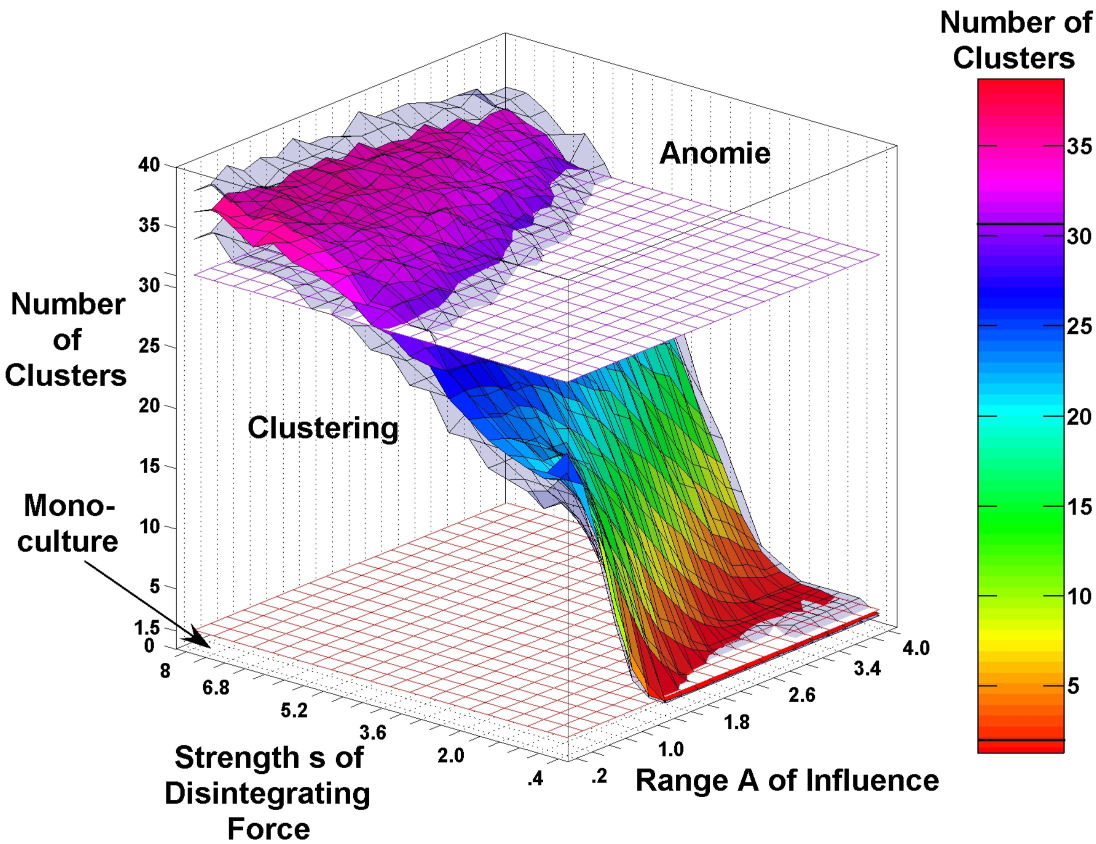


Opinion trajectories of three representative simulation runs with 100 agents over 10.000 iterations generated by the Durkheimian model. In all three runs, the opinions are restricted to values between -250 and 250, and all agents hold the same opinion initially ($o_i(0)=0$ for all i). In all runs, we assume the same social influence range $A=2$, but vary the strength s of the disintegrating force. **(A)** Monoculture, resulting in the case of a weak disintegrating force ($s=0.4$). Agents do not hold perfectly identical opinions, but the variance is low. **(B)** Anomie (i.e. extreme individualism), generated by a very strong disintegrating force ($s=6$). Agents spread over the complete opinion scale. The black line represents the time-dependent opinion of a single, randomly picked agent, showing significant opinion changes over time, which is in contrast to the collective opinion formation dynamics found in the monocultural and pluralistic cases (A) and (B). **(C)** For a moderate disintegrating force ($s=1.2$), the population quickly disintegrates into clusters. As long as these clusters are small, they are metastable. However, clusters perform random walks and can merge (e.g. around iteration 5500). As the disintegrating force grows with the size of a cluster, big clusters eventually split up into subclusters (e.g. around iteration 7000). The additional graph, in which each agent's opinion trajectory is represented by a solid black line, is an alternative visualization of the simulation run with $s=1.2$. It shows that the composition of clusters persists over long time periods.

Simulation scenarios A and B are characteristic for what Durkheim referred to as states of social cohesion and of anomie. Interestingly, however, pluralism arises as a third state in which several opinion clusters form and coexist. Figure V.2C shows a typical simulation run, where the adaptive noise maintains pluralism despite the antagonistic impacts of integrating and disintegrating forces - in fact *because* of this. In the related region of the parameter space, disintegrating forces prevent global consensus, but the integrating forces are strong enough to prevent the population from extreme individualization. This is in pronounced contrast to what we found for the BC-model with strong noise (Figure V.1C). Instead, we obtain a number of coexisting, metastable clusters of a characteristic, parameter-dependent size. Each cluster consists of a relatively small number of agents, which keeps the disintegrating forces in the cluster weak and allows clusters to persist. (Remember that the tendency of individualization according to Eq. 3 increases, when many individuals hold similar opinions.) However, due to opinion drift, distinct clusters may eventually merge. When this happens, the emergent cluster becomes unstable and will eventually split up into smaller clusters, because disintegrating forces increase in strength as a cluster grows.

Strikingly, the state of diversity, in which several opinion clusters can coexist, is not restricted to a narrow set of conditions under which integrating and disintegrating forces are balanced exactly. Figure V.3 demonstrates that opinion clusters exist in a significant area of the parameter space, i.e. the clustering state establishes another phase, which is to be distinguished from monoculture and from anomie.

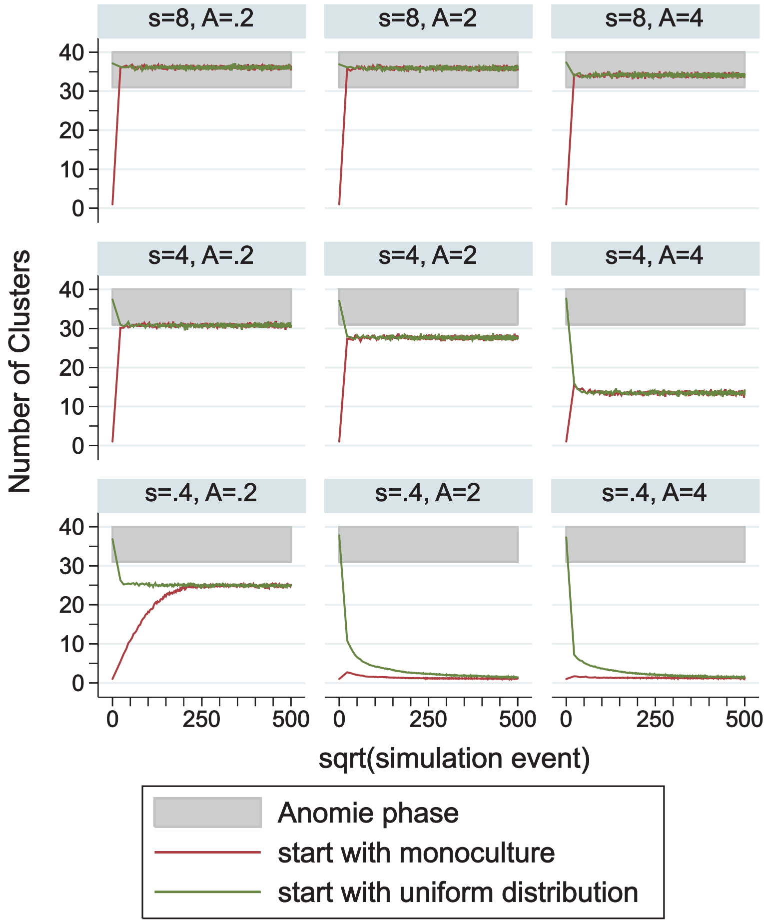
Figure V.3: Dependence of the average number of clusters in the Durkheimian model on the strength *s* of the disintegrating force and the range *A* of social influence



We conducted computer simulations with *N*=100 agents, starting with initial consensus ($o_i(0)=0$ for all *i*). We restricted opinions to values between -250 and 250. The strength *s* of the disintegrating force varied between *s*=0.4 and *s*=8 in steps of 0.4, while *A* varied between *A*=0.2 and *A*=4 in steps of 0.2. For each parameter combination, we conducted 100 independent replications and assessed the average number of clusters formed after 250,000 iterations (see *z*-axis and the color scale). The two transparent (gray) surfaces depict the inter-quartile range, which indicates a small variance in the number of clusters (and also typical cluster sizes) for each parameter combination. The horizontal grids indicate the borders of the three phases, as defined by us. An average cluster size below 1.5 indicates monoculture. Values between 1.5 and 31 reflect clustering. Finally, values above 31 correspond to opinion distributions that cannot be distinguished from random ones and represent a state of anomie.

To generate Figure V.3, we conducted a simulation experiment in which we varied the influence range *A* and the strength *s* of the disintegrating force. For each parameter combination, we ran 100 replications and measured the average number of clusters that were present after 250,000 iterations. To count the number of clusters in a population, we ordered the N agents according to their opinion. A cluster was defined as a set of agents in adjacent positions such that each set member was separated from the adjacent set members by a maximum of 5 scale points (= opinion range/N). Figure V.3 shows that, for large social influence ranges A and small noise strengths s, the average number of clusters is below 1.5, reflecting monoculture in the population. In the other extreme, i.e. for a small influence range A and large noise strengths s, the resulting distribution contains more than 31 clusters, a number of clusters that cannot be distinguished from purely random distributions. Following Durkheim, we have classified such cases as anomie, i.e. as the state of extreme individualism. Between these two phases, there are numerous parameter combinations, for which the number of clusters is higher than 1.5 and clearly smaller than in the anomie phase. This constitutes the clustering phase. Figure V.3 also shows that, for each parameter combination, there is a small variance in the number of clusters, which is due to a statistical equilibrium of occasional fusion and fission processes of opinion clusters (see Figure V.2C).

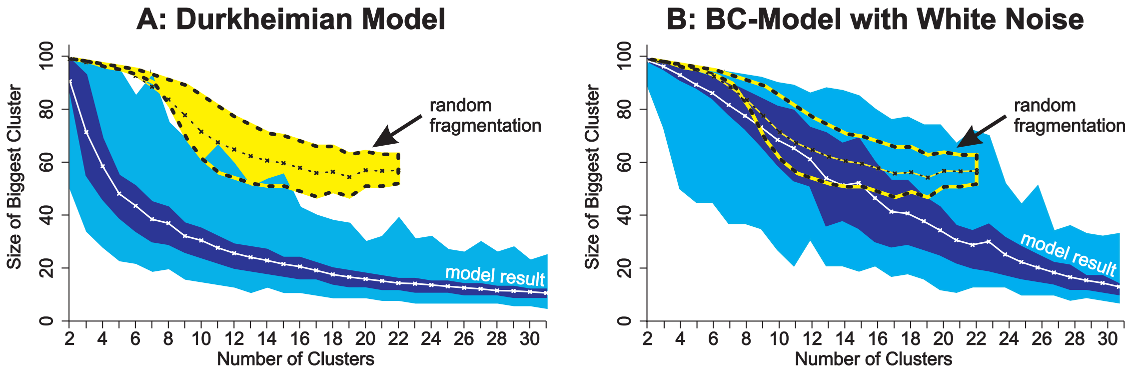
Figure V.4: Comparison of developments of average number of clusters in experiments starting with monoculture and experiments starting with uniform opinion distributions



Graphs demonstrate for 9 out of 40 conditions of the experiments that after $25*10^6$ simulation events the same average number of clusters obtained independent on the initial opinion distribution. Red lines depict average number of clusters in the experiment which started from monoculture (described in Figure V.3). The green lines show results from an identical experiment where we assigned random values (drawn from uniform distribution) to initial opinions of the agents.

The same results were found, when starting the computer simulations with a uniform opinion distribution instead of perfect monoculture. Figure V.4 compares the developments of the average the number of clusters in the main experiment with initial monoculture and the replicated experiment with uniform opinion distributions at the outset. The central message of Figure V.4 is that the trajectories of each condition approach the same value and remain stable. This clearly demonstrates that simulations ran long enough and that clustering is indeed an attractor under certain conditions.

Figure V.5: Comparison of the opinion distributions generated (A) by the Durkheimian model and (B) by the noisy BC-model



Comparison of the opinion distributions generated (A) by the Durkheimian model and (B) by the noisy BC-model, showing the size of the biggest cluster over the number of clusters (in all simulation runs that resulted in more than one and less than 32 clusters). **A** is based on the simulation experiment underlying Figure V.3. **B** was generated by a simulation experiment with the BC-model (Hegselmann and Krause 2002), in which we varied the bounded confidence level ε between 0.01 and 0.15 in steps of 0.02 and the noise level θ between 5 and 50 in steps of 5. We conducted 100 independent replications per parameter combination and measured the number of clusters and the size of the biggest cluster after 250,000 iterations.
White solid lines represent the average size of the biggest cluster. The dark blue area shows the respective interquartile range and the light blue area the complete value range of the size of the biggest cluster. For comparison, we generated randomly fragmented opinion distributions of N=100 individuals as follows: n agents were assumed to hold a random opinion, drawn from a normal distribution with an average of zero and a standard deviation of 50. The remaining N-n agents were assumed to hold opinion o_i=0, thereby forming one big cluster. We varied the value of n between 0 and 100 in steps of 1 and generated 1000 distributions per condition. For each distribution, we determined the number of clusters and the size of the biggest cluster. The average size of the biggest cluster is shown by the thin yellow-black line. (The curve stops at 22, since this is the highest number of clusters generated.) The bold yellow-black lines represent the related interquartile range.
In essence, we find that the interquartile range of the Durkheimian model is not consistent with the hypothesis of random fragmentation. Even the complete value range of the Durkheimian model (blue area) hardly overlaps with the interquartile range of the fragmented distributions (yellow area). This demonstrates that the Durkheimian model shows clustering rather than fragmentation. In contrast, Figure V. 4B illustrates that the distributions of the noisy BC-model and the results for random fragmentation overlap.

Additional statistical tests were performed to make sure that the existence of clusters in our model indeed indicates pluralism and not fragmentation, a state in which a population consists of one big cluster and a number of isolated agents (see Figure V.5). While the Durkheimian opinion dynamics model is consistent with cluster formation (see Figure V.5A), the noisy BC model rather shows random fragmentation (see Figure V.5B).

V.4. Discussion

The emergence and persistence of pluralism is a puzzling phenomenon in a world in which social networks are highly connected and social influence is an ever present force that reduces differences between those who interact. We have developed a formal theory of social influence that, besides anomie and monoculture, shows a third, pluralistic phase characterized by opinion clustering. It occurs, when all individuals interact with each other and noise prevents the convergence to a single opinion, despite homophily.

Our model does not assume negative influence, and it behaves markedly different from bounded confidence models, in which white opinion noise produces fragmentation rather than clustering. It would be natural to generalize the model in a way that also considers the structure of real social networks. This basically requires one to replace the values w_{ij} by $w_{ij}a_{ij}$, where a_{ij} are the entries of the adjacency matrix (i.e. $a_{ij} = 1$, if individuals i and j interact, otherwise $a_{ij}=0$). In such a case, resulting opinion clusters are expected to have a broad range of different sizes, similar to what is observed for the sizes of social groups.

Our model highlights the functional role that ``noise" (randomness, fluctuations, or other sources of variability) plays for the organization of social systems. It furthermore shows that the combination of two mechanisms (deterministic integrating forces and stochastic disintegrating forces) can give rise to new phenomena. We also believe that our results are meaningful for the analysis of the social integration of our societies. Both classical (Durkheim 1997 [1893]) and contemporary (Beck 1994) social thinkers argue that, in modern and globalized societies, individuals are increasingly exposed to disintegrating forces that detach them from traditional social structures. In other words, the social forces that motivate individuals to follow societal norms may lose their power to limit individual variation. Durkheim feared that this will atomize societies (Durkheim 1997 [1893]). That is, he thought the tendency towards individualization would turn societies ``anomic" as they modernize: Durkheim felt that extreme individualization in modern societies would obstruct the social structures that traditionally provided social support and guidance to individuals.

Today, modern societies are highly diverse, but at the same time they are far from a state of anomie as foreseen by Durkheim. Our model offers an explanation why and how this is possible: Besides monoculture and anomie, there is a third, pluralistic clustering phase, in which individualization prevents overall consensus, but at the same time, social influence can still prevent extreme individualism. The interplay between integrating and disintegrating forces leads to a plurality of opinions, while metastable subgroups occur, within which individuals find a local consensus. Individuals may identify with such subgroups and develop long-lasting social relationships with similar others. Therefore, they are not isolated and not without support or guidance, in contrast to the state of anomie that Durkheim was worried about.

We have seen, however, that pluralism and cultural diversity require an approximate balance between integrating and disintegrating forces. If this balance is disturbed, societies may drift towards anomie or monoculture. It is, therefore, interesting to ask how the current tendency of globalization will influence society and cultural dynamics. The Internet, interregional migration, and global tourism, for example, make it easy to get in contact with members of distant and different cultures. Previous models (Axelrod 1997; Greig 2002) suggest that this could affect cultural diversity in favor of a monoculture. However, if the individual striving for uniqueness is sufficiently strong, formation of diverse groups (a large variety of international social communities) should be able to persist even in a globalizing world. In view of the alternative futures, characterized by monoculture or pluralism, further theoretical, empirical, and experimental research should be performed to expand our knowledge of the mechanisms that will determine the future of pluralistic societies.

VI. Samenvatting - Dutch summary

Moderne samenlevingen worden gekenmerkt door een grote diversiteit van meningen. Zo tonen langetermijnstudies naar maatschappelijke houdingen van Amerikanen aan dat de meningen over belangrijke onderwerpen als gevoelens ten aanzien van Afro-Amerikanen, opvattingen over seksegelijkheid en de meningen omtrent abortus zeer gevarieerd zijn. De diversiteit van meningen is gedurende enkele tientallen jaren onveranderd zeer groot gebleven en is ten aanzien van een aantal aspecten zelfs alleen maar toegenomen. Ook in kleine sociale groepen speelt de diversiteit van meningen een belangrijke rol. Zo toont onderzoek naar de groepsdynamiek binnen organisaties aan dat teamleden er vaak niet in slagen meningsverschillen binnen het team te verkleinen.

De bestaande theorieën over de dynamiek van meningen zijn niet toereikend om dat patroon te verklaren. Die theorieën gaan uit van de veronderstelling dat mensen zo worden beïnvloed door hun netwerkcontacten dat zij steeds meer op elkaar gaan lijken. Formele modellen met betrekking tot de dynamiek van sociale beïnvloeding binnen grote populaties tonen aan dat meningen als gevolg van sociale beïnvloeding steeds meer naar elkaar toegroeien, wat uiteindelijk leidt tot volstrekte eenvormigheid van meningen.

Dit boek gaat over de tegenstelling tussen de grote diversiteit van meningen die we empirisch waarnemen en de voorspelling, op grond van modellen met betrekking tot sociale beïnvloeding, dat de diversiteit van meningen steeds verder zal afnemen. Ons doel is nieuwe theorieën te ontwikkelen die een verklaring kunnen bieden voor de diversiteit van meningen en waarmee kan worden onderzocht onder welke omstandigheden die diversiteit ontstaat. We willen vooral twee specifieke patronen van diversiteit verklaren die op basis van de bestaande theorieën niet verklaard kunnen worden: clustering van meningen en polarisatie van meningen. Met clustering van meningen wordt bedoeld het hardnekkige bestaan van verschillende subgroepen die intern homogeen zijn maar onderling afwijken. Polarisatie van meningen heeft betrekking op het ontstaan van subgroepen die geleidelijk steeds verder van elkaar verwijderde meningen ontwikkelen.

In dit boek richten we ons op drie mechanismen die we verwerken in de bestaande theorieën van sociale beïnvloeding. We ontwikkelen een aantal formele modellen waarmee we computerexperimenten uitvoeren om te testen of en onder welke omstandigheden de afzonderlijke mechanismen clustering en polarisatie kunnen helpen verklaren.

In de eerste plaats richten we onze aandacht op de aanname van *negatieve beïnvloeding*. Op basis van recente bijdragen aan de literatuur over sociale beïnvloeding gaan we ervan uit dat mensen meningsverschillen groter proberen te maken om zich te onderscheiden van anderen. Computerexperimenten tonen aan dat het nieuwe model een verklaring kan bieden voor polarisatie van meningen. Als populaties voldoende gevarieerd zijn, zullen mensen met relatief radicale meningen zich in sterke mate anders voelen dan mensen met een tegenovergestelde mening en zullen zij proberen die verschillen groter te maken door nog radicalere meningen te ontwikkelen. Dat kan invloed hebben op actoren met gematigde meningen. Ze lijken op sommige van de mensen met radicale meningen en zullen positief door hen worden beïnvloed. Ook actoren met een aanvankelijk gematigde mening zullen als gevolg daarvan radicaliseren. Het aantal mensen met een gematigde mening neemt geleidelijk af, tot de populatie is verdeeld in twee groepen die maximaal van elkaar verschillen.

Wij tonen aan dat een model waarin de aanname van negatieve beïnvloeding is inbegrepen een verklaring kan bieden voor polarisatie van meningen. We wijzen er echter ook op dat er tot dusver geen afdoend empirisch bewijs is voor negatieve beïnvloeding. Daarom ontwikkelen we een tweede theorie over polarisatie van meningen. In die benadering is het uitgangspunt dat mensen hun meningen baseren op argumenten en dat zij die argumenten onderling uitwisselen. Wanneer mensen met vergelijkbare meningen met elkaar communiceren, dan is het waarschijnlijk dat ze elkaar nieuwe argumenten aanreiken die hun meningen onderbouwen. Hun meningen worden daardoor radicaler. Wij tonen aan dat het uitwisselen van argumenten kan leiden tot polarisatie van meningen bij mensen die geneigd zijn vooral te communiceren met mensen met vergelijkbare meningen. In de interactie met gelijkgestemden wisselen mensen meestal argumenten uit die elkaars meningen zullen versterken. Actoren met enigszins gematigde opvattingen zullen daardoor geleidelijk radicaliseren. Dat proces kan zich gelijktijdig aan weerszijden van het spectrum van meningen ontwikkelen, waardoor er zich subgroepen vormen met meningen die steeds verder van elkaar verwijderd zijn.

Ons derde model is gericht op clustering van meningen. Vooraanstaande sociologische theorieën en resultaten van empirisch onderzoek suggereren dat mensen uniek willen zijn en dat zij hun meningen aanpassen wanneer zij merken dat er te veel anderen zijn die dezelfde mening als hen hebben. Wij passen dat idee toe in de structuur van sociale beïnvloeding en tonen aan dat het ontstaan van verschillende homogene clusters met uiteenlopende meningen daardoor kan worden verklaard. Clusters zijn, wanneer ze eenmaal zijn ontstaan, tijdelijk stabiel maar kunnen wel samensmelten. Als dat gebeurt, kan dat weer het ontstaan van nieuwe clusters veroorzaken. Zo is clustering een stabiel resultaat.

Op basis van de nieuwe theorieën over diversiteit van meningen onderzoek wij onder welke omstandigheden meningen polariseren. We richten onze aandacht vooral op de effecten van demografische diversiteit op de dynamiek van meningen binnen werkteams. Onderzoek toont aan dat diversiteit op het gebied van aspecten als etnische achtergrond, religie en geslacht een gunstig effect op teams kan hebben omdat het sociale en cultureele kapitaal waarmee teams hun taken kunnen volbrengen, wordt uitgebreid. Tegelijkertijd kunnen demografische verschillen echter ook aanleiding geven tot meningsverschillen en spanningen die de prestatie in gevaar brengen. Deze overwegingen geven aanleiding tot de heel praktische vraag onder welke omstandigheden demografisch gevarieerde werkteams erin zullen slagen meningsverschillen te overwinnen en in staat zullen zijn gebruik te maken van de voordelen die zij bieden.

Onze modellen bekrachtigen het idee dat demografische diversiteit dynamiek veroorzaakt in meningen in werkteams. We tonen echter aan dat ook zeer gevarieerde teams meningsverschillen kunnen overwinnen. We tonen vooral aan dat het effect van demografische verscheidenheid op de polarisatie van meningen afhankelijk is van de mate waarin er onderlinge samenhang bestaat tussen demografische eigenschappen in de groep. Zo is een groep bestaande uit twee zwarte mannen en twee blanke vrouwen vatbaar voor polarisatie van meningen. Die groep bestaat immers uit twee demografisch homogene maar onderling zeer verschillende subgroepen. Die verschillen zouden kunnen leiden tot negatieve beïnvloeding, wat kan leiden tot polarisatie. Een groep bestaande uit een zwarte en een blanke vrouw en een zwarte en een blanke man heeft dezelfde demografische verscheidenheid maar er is geen onderlinge relatie tussen de demografische eigenschappen. De teamleden hebben dus met bijna al hun collega's een aantal demografische eigenschappen gemeen. Die overeenkomsten kunnen negatieve beïnvloeding voorkomen en zo polarisatie afremmen.

Om kort te gaan: dit boek bespreekt nieuwe oplossingen voor het tot dusver onopgelost gebleven vraagstuk van het bestaan van diversiteit van meningen ondanks sociale beïnvloeding. Er is nog veel werk te doen, maar wij denken dat de theorieën die in dit boek naar voren worden gebracht nieuwe uitgangspunten bieden voor vruchtbaar nader onderzoek. Ze reiken bovendien mogelijk nieuwe benaderingen aan om om te gaan met de dynamiek van meningen in groepen die beroepsbeoefenaren kunnen helpen polarisatie te voorkomen of, indien nodig, een gezonde diversiteit van meningen aan te moedigen.

VII. References

Abelson, Robert P. 1964. "Mathematical Models of the Distribution of Attitudes Under Controversy." Pp. 142-160 in *Contributions to Mathematical Psychology*, edited by N. Frederiksen and H. Gulliksen. New York: Rinehart Winston.

Abramowitz, A. I., and K. L. Saunders. 2008. "Is polarization a myth?" *Journal of Politics* 70:542-555.

Adner, Ron, Laszlo Polos, Michael Ryall, and Olav Sorenson. 2009. "Introduction to Special Topic Forum. The Case for Formal Theory." *Academy of Management Review* 34:201-208.

Akers, Ronald L., Marvin D. Krohn, Lonn Lanza-Kaduce, and Marcia Radosevich. 1979. "Social Learning and Deviant Behavior: A specific Test of a general Theory." *American Sociological Review* 44:636-655.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Boston: Beacon Press.

Anderson, Norman H. 1991. *Contributions to Information Integration Theory*. Hillsdale, NJ.: Lawrence Erlbaum.

Aral, S., Lev Muchnik, and Arun Sundararajan. 2009. "Distinguishing influenced-based contagion from homophily-driven diffusion in dynamic networks." *Proceedings of the National Academy of Sciences of the United States of America* 106:21544-21549.

Axelrod, R. 1997. "The dissemination of culture - A model with local convergence and global polarization." *Journal of Conflict Resolution* 41:203-226.

Bacharach, Samuel B., Peter A. Bamberger, and Dana Vashdi. 2005. "Diversity and Homophily at work: Supportive Relations among White and African-American Peers." *Academy of Management Journal* 48:619-644.

Baldassarri, D., and P. Bearman. 2007. "Dynamics of political polarization." *American Sociological Review* 72:784-811.

Barabasi, A. L. 2003. *Linked: How Everything Is Connected to Everything Else and What It Means*. New York: Plume.

Barabasi, A. L., and R. Albert. 1999. "Emergence of scaling in random networks." *Science* 286:509-512.

Beck, Ulrich. 1994. "The Reinvention of Politics: Towards a Theory of Reflexive Modernization." Pp. 1-55 in *Reflexive Modernization. Politics, Tradition and Aestherics in the Modern Social Order*, edited by Ulrich Beck, Anthony Giddens, and Scott Lash. Cambridge: Polity Press.

Berger, J., and C. Heath. 2008. "Who drives divergence? Identity signaling, outgroup dissimilarity, and the abandonment of cultural tastes." *Journal of Personality and Social Psychology* 95:593-607.

Berger, Roger L. 1981. "A Necessary and Sufficient Condition for Reaching a Consensus Using DeGroot's Method." *Journal of the American Statistical Association* 76:415-418.

Berscheid, Ellen. 1966. "Opinion Change and Communicator-Communicatee Similarity and Dissimilarity." *Journal of Personality and Social Psychology* 4:670-680.

Bezrukova, K., K. A. Jehn, E. L. Zanutto, and S. M. B. Thatcher. 2009. "Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance." *Organization Science* 20:35-50.

Bezrukova, Katerina, S. B. M. Thatcher, and Karen Jehn. 2007. "Group heterogeneity and faultlines: Comparing alignment and dispersion theories of group composition." Pp. 57-92 in *Conflict in Organizational Groups: New Directions in Theory and Practice*, edited by L. Thompson and K. Behfar. Evanston, IL: Northwestern University Press.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural-Change as Informational Cascades." *Journal of Political Economy* 100:992-1026.

Bonabeau, E. 2002. "Agent-based modeling: Methods and techniques for simulating human systems." *Proceedings of the National Academy of Sciences of the United States of America* 99:7280-7287.

Bourdieu, Pierre. 1984[1979]. *Distinction: A social critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.

Bowers, Clint A., James A. Pharmer, and Eduardo Salas. 2000. "When Member Homogeneity is needed in Work Teams. A Meta-Analysis." *Small Group Research* 31:305-327.

Brass, Daniel J., Joseph Galaskiewicz, Henrich R. Greve, and Wenpin Tsai. 2004. "Taking Stock of Networks and Organizations: A Multilevel Perspective." *Academy of Management Journal* 47:795-817.

Brewer, M. B. 1991. "The Social Self - on Being the Same and Different at the Same Time." *Personality and Social Psychology Bulletin* 17:475-482.

Brown, G. D. A., and N. Chater. 2001. "The Chronological Organization of Memory: Common Psychological Foundations for Remembering and Timing." in *Time and memory: Issues in Philosophy and Psychology*, edited by Christoph Hoerl and Teresa McCormack. Oxford: Oxford University Press.

Bryson, B. 1996. ""Anything but heavy metal": Symbolic exclusion and musical dislikes." *American Sociological Review* 61:884-899.

Burk, W. J., C. E. G. Steglich, and A. B. Snijders. 2007. "Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors." *International Journal of Behavioral Development* 31:397-404.

Byrne, D. 1961. "Interpersonal-Attraction and Attitude Similarity." *Journal of Abnormal and Social Psychology* 62:713-715.

Byrne, Donn. 1971. *The Attraction Paradigm*. New York, London: Academic Press.

Byrne, Donn, Gerald L. Clore, and George Smeaton. 1986. "The Attraction Hypothesis: Do Similar Attitudes Affect Anything?" *Journal of Personality and Social Psychology* 51:1167-1170.

Carley, Kathleen. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331-354.

Castellano, C., S. Fortunato, and V. Loreto. 2009. "Statistical physics of social dynamics." *Reviews of Modern Physics* 81:591-646.

Chatman, Jennifer A., Jeffrey T. Polzer, Sigal G. Barsade, and Margaret A Neale. 1998. "Being Different Yet Feeling Similar: The Influence of Demographic Composition and Organizational Culture on Work Process and Outcome." *Administrative Science Quarterly* 43:749-780.

Chen, Fang Fang, and Douglas T. Kenrick. 2002. "Repulsion or Attraction? Group Membership and Assumed Attitude Similarity." *Journal of Personality and Social Psychology* 83:111-125.

Colson, Elizabeth. 1954. "Social Control and Vengeance in Plateau Tonga Society." *Africa: Journal of the International African Institute* 23:199-212.

Cowan, N. 2001. "The magical number 4 in short-term memory: A reconsideration of mental storage capacity." *Behavioral and Brain Sciences* 24:87-185.

Deffuant, G., Sylvie Huet, and Frédéric Amblard. 2005. "An Individual-Based Model of Innovation Diffusion Mixing Social Value and Individual Benefit." *American Journal of Sociology* 110:1041-1069.

DeGroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69:118-121.

DiMaggio, P., J. Evans, and B. Bryson. 1996. "Have Americans' social attitudes become more polarized?" *American Journal of Sociology* 102:690-755.

Durkheim, Emile. 1973. *On Morality and Society. Selected Writings*. Chicago: University Of Chicago Press.

—. 1982 [1895]. *The Rules of Sociological Method*. New York: The Free Press.

—. 1997 [1893]. *The Division of Labor in Society*. New York: The Free Press.

Early, P. Christopher, and Elaine Mosakowski. 2000. "Creating Hybrid Team Cultures: An Empirical Test of Transnational Team Functioning." *Academy of Management Journal* 43:26-49.

Elias, Norbert. 1969[1939]. *The Civilizing Process, Vol.I. The History of Manners*. Oxford: Blackwell.

Evans-Pritchard, E. E. 1939. "Introduction." Pp. XIX-XXXIV in *The Social institutions of the Kipsigis*, edited by J.G. Peristiany. London: Routledge.

Evans, J. 2003. "Have Americans' Attitudes Become More Polarized?-An Update." *Social Science Quarterly* 84:71-90.

Feldman, K. A., and T. M. Newcomb. 1969. *The impact of college on students*. San Francisco: Jossey-Bass.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Evanston, White Plains: Row, Petersen and Company.

Festinger, Leon, S. Schachter, and K. Back. 1950. *Social Pressures in Informal Groups*. Stanford, CA: Stanford University Press.

Fiorina, M. P., and S. J. Abrams. 2008. "Political polarization in the American public." *Annual Review of Political Science* 11:563-588.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10:171-178.

Fischer, Claude S., and Greggor Mattson. 2009. "Is America Fragmenting?" *Annual Review of Sociology* 35:435-455.

Flache, A., and M. Macy. 2006a. "What sustains cultural diversity and what undermines it? Axelrod and beyond." arXiv:physics/0604201v1 [physics.soc-ph].

Flache, Andreas, and Michael W. Macy. 2006b. "Why more contact may increase cultural polarization." in *Session Mathematical Sociology I at 101st Annual Meeting of the American Sociological Association*. Montreal.

Flache, Andreas, Michael W. Macy, and Károly Takács. 2006. "What Sustains Stable Cultural Diversity and What undermines it? Axelrod and Beyond." in *World Congress of Social Simulation 2006*. Kyoto, Japan.

Flache, Andreas, and Michael Mäs. 2008a. "How to get the timing right? A computational model of how demographic faultlines undermine team performance and how the right timing of contacts can solve the problem." *Computational and Mathematical Organization Theory* 14:23-51.

—. 2008b. "Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion." *Simulation Modelling Practice and Theory* 16:175-191.

Flap, Henk. 1988. *Conflict, loyalty and violence: The effects of social networks on behaviour*. Frankfurt am Main: Peter Lang.

Frank, K. A., and K. Fahrbach. 1999. "Organization culture as a complex system: Balance and information in models of influence and selection." *Organization Science* 10:253-277.

French, J. R. P. 1956. "A Formal Theory of Social Power." *Psychological Review* 63:181-194.

Friedkin, N. E., and E. C. Johnsen. 1990. "Social-Influence and Opinions." *Journal of Mathematical Sociology* 15:193-205.

—. 1997. "Social positions in influence networks." *Social Networks* 19:209-222.

Friedman, Thomas L. 2005. *The World is Flat. A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.

Galam, S. 2005. "Heterogeneous beliefs, segregation, and extremism in the making of public opinions." *Physical Review E* 71:-.

Galtung, Johan. 1966. "Rank and Social Integration: A Multidimensional Appraoch." in *Sociological Theories in Progress*, edited by Joseph Berger, Morris Zelditch, and Bo Anderson. Boston: Houghton Mifflin Company.

Gibson, Cristina, and Freek Vermeulen. 2003. "A Healthy Divide: Subgroups as a Stimulus for Team Learning Behavior." *Administrative Science Quarterly* 48:202-239.

Gibson, David R. 2005. "Concurrency and Commitment: Network Scheduling and Its Consequences for Diffusion." *Journal of Mathematical Sociology* 29:295-323.

Glaeser, E. L., and B. A. Ward. 2006. "Myths and realities of American political geography." *Journal of Economic Perspectives* 20:119-144.

Greig, Michael J. 2002. "The End of Geography? Globalization, Communication and Culture in the international System." *Journal of Conflict Resolution* 46:225-243.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." Pp. 79-93 in *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, edited by Kurt Kremer and Volker Macho. Göttingen.

Harary, F. 1959. "A criterion for unanimity in French's theory of social power." Pp. 168-182 in *Studies in social power*, edited by D. Cartwright. Ann Arbor: Institute for Social Research.

Harrison, David A., Kenneth H. Price, Joanne H. Gavin, and Anna T. Florey. 2002. "Time, Teams, and Task Performance: Changing effects of surface and deep-level diversity on group functioning." *Academy of Management Journal* 45:1029-1045.

Harrison, J. R., Z. Lin, G. R. Carroll, and K. M. Carley. 2007. "Simulation modeling in organizational and management research." *Academy of Management Review* 32:1229-1245.

Harrison, J.R., and G. Carroll. 2002. "The Dynamics of Cultural Influence Networks." *Computational and Mathematical Organization Theory* 8:5-30.

Hart, C. M., and M. Van Vugt. 2006. "From fault line to group fission: Understanding membership changes in small groups." *Personality and Social Psychology Bulletin* 32:392-404.

Hegselmann, Rainer, Andreas Flache, and Volker Möller. 2002. "Cellular Automata as a Modelling Tool: Solidarity and Opinion Formation." Pp. 151-178 in *Tools and Techniques for the Social Sciences Simulation*, edited by Suleiman R., Klaus G. Troitzsch, and Nigel Gilbert. Heidelberg: Physica.

Hegselmann, Rainer, and Ulrich Krause. 2002. "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation." *Journal of Artificial Societies and Social Simulation* 5.

Heider, Fritz. 1967. "Attitudes and Cognitive Organization." Pp. 39-41 in *Readings in Attitude Theory and Measurement*, edited by Martin Fishbein. New York, London, Sydney: John Wiley and Sons, Inc.

Hogg, M. A., J. C. Turner, and B. Davidson. 1990. "Polarized Norms and Social Frames of Reference: A Test of the Self-Categorization Theory of Group Polarization." *Basic and Applied Social Psychology* 11:77-100.

Homan, A. C., D. van Knippenberg, G. A. Van Kleef, and C. K. W. De Dreu. 2007. "Interacting dimensions of diversity: Cross-categorization and the functioning of diverse work groups." *Group Dynamics-Theory Research and Practice* 11:79-94.

Homans, George C. 1951. *The Human Group*. New York: Harcourt Press.

Hornsey, Matthew J., and Jolanda Jetten. 2004. "The Individual Within the Group: Balancing the Need to Belong With the Need to be Different." *Personality and Social Psychology Review* 8:248-264.

Ibarra, H. 1992. "Homophily and Differential Returns - Sex-Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37:422-447.

Imhoff, R., and H. P. Erb. 2009. "What Motivates Nonconformity? Uniqueness Seeking Blocks Majority Influence." *Personality and Social Psychology Bulletin* 35:309-320.

Isenberg, Daniel J. 1986. "Group Polarization: A Critical Review and Meta-Analysis." *Journal of Personality and Social Psychology* 50:1141-1151.

Jager, Wander, and Frédéric Amblard. 2005. "Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change." *Computational and Mathematical Organization Theory* 10:295-303.

Jehn, Karen A. 1994. "Enhancing Effectiveness: An Investigation of Advantages and Disadvantages of Value-Based Intragroup Conflict." *The International Journal of Conflict Management* 5:223-238.

—. 1995. "A Multimethod Examination of the Benefits and Detriments of Intragroup Conflict." *Administrative Science Quarterly* 40:256-282.

—. 1997. "A Qualitative Analysis of Conflict Types and Dimensions of Organizational Groups." *Administrative Science Quarterly* 42:530-557.

Jehn, Karen A., and Katerina Bezrukova. 2004. "A Field Study of Group Diversity, Workgroup context, and Performance." *Journal of Organizational Behavior* 25:703-729.

Jehn, Karen A., Gregory B. Northcraft, and Margaret A Neale. 1999. "Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups." *Administrative Science Quarterly* 44:741-763.

Johnson, D. W., and F. P. Johnson. 1982. *Joining Together: Group theory and skills*. Englewood Cliffs, NJ: Prentice Hall.

Kamada, T., and S. Kawai. 1989. "An Algorithm for drawing general undirected graphs." *Information Processing Letters* 31:7-15.

Kameda, Tatsuya, and Shinkichi Sugimori. 1995. "Procedural Influence in Two-Step Group Decision Making: Power of Local Majorities in Consensus Formation." *Journal of Personality and Social Psychology* 69:865-876.

Kandel, Denise B. 1978. "Homophily, Selction, and Socialization in Adolescent Friendships." *American Journal of Sociology* 84:427-436.

Kennedy, James. 1998. "Thinking is Social." *Journal of Conflict Resolution* 42:56-76.

Kerr, Norbert L., and Scott R. Tindale. 2004. "Group Performance and Decision Making." *Annual Review of Psychology* 55:623-655.

Kitts, James. 2006. "Social Influence and the Emergence of Norms Amid Ties of Amity and Enmity." *Simulation Modelling Practice and Theory* 14:407-422.

Klemm, K., V. M. Eguiluz, R. Toral, and M. S. Miguel. 2003a. "Global culture: A noise-induced transition in finite systems." *Physical Review E* 67:045101(R).

Klemm, K., V. M. Eguiluz, R. Toral, and M. San Miguel. 2003b. "Nonequilibrium transitions in complex networks: A model of social interaction." *Physical Review E* 67:026120 (R).

Krizan, Z., and R. S. Baron. 2007. "Group polarization and choice-dilemmas: How important is self-categorization?" *European Journal of Social Psychology* 37:191-201.

Lau, Dora C., and J. Keith Murnighan. 1998. "Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups." *Academy of Management Review* 23:325-340.

—. 2005. "Interactions within Groups and Subgroups: The Effects of Demographic Faultlines." *Academy of Management Journal* 48:645-659.

Lazarsfeld, Paul F., and Robert K. Merton. 1954. "Friendship and Social Process: A Substantive and Methodological Analysis." Pp. 18-66 in *Freedom and Control in Modern Society*, edited by Morroe Berger, Theordore Abel, and Charles H. Page. New York, Toronto, London: Van Nostrand.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. 2009. "Computational Social Science." *Science* 323:721-723.

Lehrer, K. 1975. "Social Consensus and rational agnoiology." *Synthese* 31:141-160.

Lemaine, G. 1975. "Dissimilation and Differential Assimilation in Social Influence (Situations of Normalization)." *European Journal of Social Psychology* 5:93-120.

Levendusky, Matthew S. 2009. "The Microfoundations of Mass Polarization." *Political Analysis* 17:162-176.

Levine, Michael E., and Charles R. Plott. 1977. "Agenda Influence and Its Implications." *Virginia Law Review* 63:561-604.

Li, Jiatao, and Donald C. Hambrick. 2005. "Factional Groups: A New Vantage on Demographic Faultlines, Conflict, and Disintegration in Work Teams." *Academy of Management Journal* 48:794-813.

Liben-Nowell, D., and J. Kleinberg. 2008. "Tracing information flow on a global scale using Internet chain-letter data." *Proceedings of the National Academy of Sciences of the United States of America* 105:4633-4638.

Liggett, Thomas M. 1985. *Interacting Particle Systems*. New York: Springer.

Lijphart, A. 1977. *Democracy in Plural Societies. A Comparative Exploration*. New Haven: Yale University Press.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg. 2001. "The web of human sexual contacts." *Nature* 411:907-908.

Lindenberg, Siegwart. 1992. "The Method of Decreasing Abstraction." Pp. 3-20 in *Rational Choice Theory: Advocacy and Critique*, edited by James S. Coleman and Feraro T.J. Newbury Park: Sage.

List, Christian. 2004. "A Model of Path-Dependence in Decisions over Multiple Propositions." *American Political Science Review* 98:495-513.

Lorenz, J. 2005. "A stabilization theorem for dynamics of continuous opinions." *Physica a-Statistical Mechanics and Its Applications* 355:217-223.

—. 2007. "Continuous opinion dynamics under bounded confidence: A survey." *International Journal of Modern Physics C* 18:1819-1838.

Mackie, D. M. 1986. "Social Identification Effects in Group Polarization." *Journal of Personality and Social Psychology* 50:720-728.

Macy, M. W., James Kitts, Andreas Flache, and Steve Benard. 2003. "Polarization and Dynamic Networks. A Hopfield Model of Emergent Structure." Pp. 162-173 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by Ronald Breiger, Kathleen Carley, and Philippa Pattison. Washington, DC: The National Academies Press.

Macy, M. W., and R. Willer. 2002. "From factors to actors: Computational sociology and agent-based modeling." *Annual Review of Sociology* 28:143-166.

Macy, Michael, and Andreas Flache. 2009. "Social Dynamics from the Bottom up. Agent-Based Models of Social Interaction." Pp. 245-268 in *The Oxford Handbook of Analytical Sociology*, edited by Peter Hedström and Peter Bearman. Oxford: Oxford University Press.

Mark, Noah P. 1998. "Beyond individual differences: Social differentiation from first principles." *American Sociological Review* 63:309-330.

—. 2003. "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches." *American Sociological Review* 68:319-345.

Mäs, Michael, A. Flache, Károly Takács, and Karen Jehn. 2008. "In the short term we divide, in the long term we unite. Crisscrossing work team members and the effects of faultlines on intergroup polarization." Groningen.

Maslach, C., J. Stapp, and R. T. Santee. 1985. "Individuation - Conceptual Analysis and Assessment." *Journal of Personality and Social Psychology* 49:729-738.

Mason, Claire M. 2006. "Exploring the Process Underlying Within-Group Homogeneity." *Small Group Research* 37:233-270.

Mason, W. A., F. R. Conrey, and E. R. Smith. 2007. "Situating social influence processes: Dynamic, multidirectional flows of influence within social networks." *Personality and Social Psychology Review* 11:279-300.

Mazen, Ragaa, and Howard Leventhal. 1972. "The Influence of Communicator-Recipient Similarity upon the beliefs and behavior of pregnant woman." *Journal of Experimental Social Psychology* 8:289-302.

McFarland, Dan, and Skye Bender-deMoll. 2007. "SoNIA (Social Network Image Animator)."

McPherson, M., L. Smith-Lovin, and M. E. Brashears. 2006. "Social isolation in America: Changes in core discussion networks over two decades." *American Sociological Review* 71:353-375.

McPherson, Miller, and L. Smith-Lovin. 1987. "Homophily in Voluntary Organizations - Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52:370-379.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-444.

Mercken, L., T. A. B. Snijders, C. E. G. Steglich, E. Vartiainen, and H. de Vries. forthcoming. "Dynamics of adolescent friendship networks and smoking behavior." *Social Networks*.

Milgram, S. 1967. "Small-World Problem." *Psychology Today* 1:61-67.

Miller, G. A. 1956. "The Magical Number 7, Plus or Minus 2 - Some Limits on Our Capacity for Processing Information." *Psychological Review* 63:81-97.

Milliken, Frances J., and Luis L. Martins. 1996. "Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups." *Academy of Management Journal* 21:402-433.

Molleman, Eric. 2005. "Diversity in Demographic Characteristics, Abilities and Personality Traits: Do Faultlines Affect Team Functioning?" *Group Decision and Negotiation* 14:173-193.

Molleman, Eric, and Jannes Slomp. 2006. "The Impact of Team and Work Characteristics on Team Functioning." *Human Factors and Ergonomics in Manufacturing* 16:1-15.

Moody, James. 2001. "Race, school integration, and friendship segregation in America." *American Journal of Sociology* 107:679-716.

—. 2002. "The Imortance of Relationship Timing for Diffusion." *Social Forces* 81:25-56.

Mouw, Ted, and Michael E. Sobel. 2001. "Culture Wars and Opinion Polarization: The Case of Abortion." *American Journal of Sociology* 106:913-943.

Mummendey, Amélie, Andreas Klink, Rosemarie Mielke, Michael Wenzel, and Mathias Blanz. 1999. "Socio-structural charcteristics of intergroup relations and identity management strategies: Results from a field study in East Germany." *European Journal of Social Psychology* 29:259-285.

Myers, D. G. 1982. "Polarizing Effects of Social Interaction." Pp. 125-161 in *Group Decision Making*, edited by Hermann Brandstätter, J. H. Davis, and Gisela Stocker-Kreichgauer. London: Academic Press.

Myers, D. G., and G. D. Bishop. 1970. "Discussion Effects on Racial Attitudes." *Science* 169:778-779.

Myers, D. G., and H. Lamm. 1976. "The Group Polarization Phenomenon." *Psychological Bulletin* 83:602-627.

Newman, M. E. J. 2004. "Colloquium Paper: Mapping Knowledge Domains: Coauthorship networks and patterns of scientific collaboration." *Proceedings of the National Academy of Sciences of the United States of America* 101:5200-5205.

Nieuwbeerta, P., and H. Flap. 2000. "Crosscutting social circles and political choice - Effects of personal network composition on voting behavior in The Netherlands." *Social Networks* 22:313-335.

Nowak, Andrzej, Jacek Szamrej, and Bibb Latané. 1990. "From Private Attitude to Public Opinion: A Dynamic Theory of Social Impact." *Psychological Review* 97:362-376.

Palla, G., A. L. Barabasi, and T. Vicsek. 2007. "Quantifying social group evolution." *Nature* 446:664-667.

Pearson, Michael, C.E.G. Steglich, and T.A.B. Snijders. 2006. "Homophily and assimilation among sportactive adolescent substance users." *Connections* 27:47-63.

Pelled, Lisa Hope. 1996. "Demographic Diversity, Conflict, and Work Group Outcomes: An Intervening Process Theory." *Organization Science* 7:615-631.

Pelled, Lisa Hope, Kathleen M. Eisenhardt, and Katherine R. Xin. 1999. "Exploring the Black Box: An Analysis of Work Group Diversity, Conflict and Performance." *Administrative Science Quarterly* 44:1-28.

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49:65-85.

Pfeffer, Jeffrey. 1985. "Organizational Demography: Implications for Management." *California Management Review* 28:67-81.

Phillips, K. W. 2003. "The Effects of categorically based Expectations on Minority Influence: The Importance of Congruence." *Personality and Social Psychology Bulletin* 29:3-13.

Phillips, K. W., E. A. Mannix, M. A. Neale, and D. H. Gruenfeld. 2004. "Diverse groups and information sharing: The effects of congruent ties." *Journal of Experimental Social Psychology* 40:497-510.

Pilkington, Neil W., and John E. Lydon. 1997. "The Relative Effect of Attitude Similarity and Attitude Dissimilarity on Interpersonal Attraction: Investigating the Moderating Roles of Prejudice and Group Membership." *Personality and Social Psychology Bulletin* 23:107-116.

Pineda, M., R. Toral, and E. Hernandez-Garcıa. 2009. "Noisy continuous-opinion dynamics." *Journal of Statistical Mechanics* P08001.

Plott, Charles R., and Michael E. Levine. 1978. "A Model of Agenda Influence on Committee Decisions." *American Economic Review* 68:146-160.

Popper, Karl. 1972. *Objective Knowledge. An Evolutionary Approach*. Oxford: The Clarendon Press.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Rainio, Kullervo. 1961a. *A Stochastic Model of Social Interaction*. Turku: Munksgaard.

—. 1961b. "Stochastic Process of Social Interaction." *Scandinavian Journal of Psychology* 2:113-128.

—. 1962. *A Stochastic Theory of Social Contacts. A Laboratory Study and an Application to Sociometry*. Turku: Munksgaard.

—. 1965. "Social Interaction as a Stochastic Learning Process." *Archives Europeennes de Sociologie* 6:68-88.

Reagans, Ray, and Ezra W. Zuckerman. 2001. "Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams." *Organization Science* 12:502-517.

Rico, Ramón, Eric Molleman, Miriam Sánchez-Manzanares, and Gerben Van der Vegt. 2007. "The effects of diversity faultlines and team task autonomy on decision quality and social integration." *Journal of Management* 33:111-132.

Rogers, Everett M., and Dilip K. Bhowmik. 1970. "Homophily-Heterophily: Relational Concepts for Communication Research." *Public Opinion Quarterly* 34:523-538.

Rosenbaum, Milton E. 1986a. "Comment on a Proposed Two-Stage Theory of Relationship Formation: First, Repulsion; Then, Attraction." *Journal of Personality and Social Psychology* 51:1171-1172.

—. 1986b. "The Repulsion Hypothesis: On the Nondevelopment of Relationships." *Journal of Personality and Social Psychology* 51:1156-1166.

Ross, E.A. 1920. *The Principles of Sociology*. New York: Century Co.

Ruef, M., H. E. Aldrich, and N. M. Carter. 2003. "The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneur." *American Sociological Review* 68:195-222.

Salzarulo, Laurent. 2006. "A Continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast." *Journal of Artificial Societies and Social Simulation* 9.

Sampson, Edward E., and Chester A. Insko. 1964. "Cognitive consistency and performance in the autokinetic situation." *Journal of Abnormal and Social Psychology* 68:184-192.

Schwartz, Shalom H., and Ruth E. Ames. 1977. "Positive and Negative Referent Others as Sources of Influence: A Case of Helping." *Sociometry* 40:12-21.

Sherif, Muzafer. 1966. *Group Conflict and Cooperation. Their Social Psychology*. London: Routlage & Kegan Paul.

Siggelkow, N., and J. W. Rivkin. 2006. "When exploration backfires: Unintended consequences of multilevel organizational search." *Academy of Management Journal* 49:779-795.

Simmel, G. 1957. "Fashion." *American Journal of Sociology* 62:541-558.

Simmel, Georg. 1922 (1908). *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. München: Duncker & Humbold.

—. 2004 (1858). *Über soziale Differenzierung*. http://www.digbib.org/Georg_Simmel_1858/Ueber_sociale_Differenzierung.

Smeaton, George, Donn Byrne, and Sarah K. Murnen. 1989. "The Repulsion Hypothesis Revisited: Similarity Irrelevance or Dissimilarity Bias?" *Journal of Personality and Social Psychology* 56:54-59.

Smith, E. R., and F. R. Conrey. 2007. "Agent-based modeling: A new approach for theory building in social psychology." *Personality and Social Psychology Review* 11:87-104.

Snijders, T.A.B., and Roel Bosker. 1999. *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Snyder, C. R., and Howard L. Fromkin. 1980. *Uniqueness. The human Pursuit of Difference*. New York and London: Plenum Press.

Stanley, H. E. 1971. *Introduction to Phase Transitions and Critical Phenomena*. Oxford and New York: Oxford University Press.

Stasser, G. 1988. "Computer-Simulation as a Research Tool - the Discuss Model of Group Decision-Making." *Journal of Experimental Social Psychology* 24:393-422.

Stauffer, D., and S. Solomon. 2007. "Ising, Schelling and self-organising segregation." *European Physical Journal B* 57:473-479.

Stewart, Greg L. 2006. "A Meta-Analytic Review of Relationships Between Team Design Features and Team Performance." *Journal of Management* 32:29-54.

Stokman, F. N., and E. P. H. Zeggelink. 1996a. "Is politics power or policy oriented? A comparative analysis of dynamic access models in policy networks." *Journal of Mathematical Sociology* 21:77-111.

Stokman, Frans, and Evelien Zeggelink. 1996b. "Self-organizing Friendship Networks." Pp. 385-418 in *Frontiers in Social Dilemmas Research*, edited by Wim B.G. Liebrand and David M. Messick. Berlin: Springer.

Sunstein, C. R. 2008. "Enclave Extremism and Journalism's Brave New World." *Nieman Reports*:34-35.

Surowiecki, James. 2005. *The wisdom of the crowds*. New York: Anchor.

Sutherland, W. J. 2003. "Parallel extinction risk and global distribution of languages and species." *Nature* 423:276-279.

Sznajd-Weron, K, and J Sznajd. 2000. "Opinion Evolution in Closed Community." *International Journal of Modern Physics C* 11:1157-1165.

Tajfel, Henri. 1981. *Human groups and social categories. Studies in social psychology*. Cambridge, London, New York, New Rochelle, Melbourne and Syndey: Cambridge University Press.

Tajfel, Henri, and John C. Turner. 1986. "The Social Identity Theory of Intergroup Behavior." Pp. 7-24 in *Psychology of Intergroup Relations*, edited by Stephen Worchel and William G. Austin. Chicago: Nelson-Hall Publishers.

Tea Party Patriots, Inc. 2010. www.teapartypatriots.org, March 20, 2010

Thatcher, Sherry M.B., Karen A. Jehn, and Elaine Zanutto. 2003. "Cracks in Diversity Research: The Effects of Diversity Faultlines on Contact and Performance." *Group Decision and Negotiation* 12:217-241.

The Coffee Party USA. 2010. www.coffeepartyusa.com, March 30, 2010

Travers, J., and S. Milgram. 1969. "Experimental Study of Small World Problem." *Sociometry* 32:425-443.

Tsuji, Ryuhei. 2002. "Interpersonal Influence and Attitude Change Toward Conformity in Small Groups: A Social Psychological Model." *Journal of Mathematical Sociology* 26:17-34.

Turner, John C. 1987. "A Self-Categorization Theory." Pp. 42 - 67 in *Rediscovering the Social Group: A Self-Categorization Theory*, edited by John C. Turner, Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell. Oxford: Blackwell.

Turner, Jonathan H. 1995. *Macrodynamics. Toward a theory on the organization of human populations*. New Brunswick, N.J: Rutgers University Press.

van Knippenberg, Ad, and H. Wilke. 1988. "Social Categorization and Attitude-Change." *European Journal of Social Psychology* 18:395-406.

van Knippenberg, D., N. De Vries, and A. van Knippenberg. 1990. "Group Status, Group-Size and Attitude Polarization." *European Journal of Social Psychology* 20:253-257.

van Knippenberg, Daan, Carsten K. W. De Dreu, and Astrid C. Homan. 2004. "Work Group Diversity and group Performance: An Integrative Model and Research Agenda." *Journal of Applied Psychology* 89:1008-1022.

van Knippenberg, Daan, and Michaéla C. Schippers. 2007. "Work Group Diversity." *Annual Review of Psychology* 58:515-541.

Vignoles, V. L., X. Chryssochoou, and G. M. Breakwell. 2000. "The distinctiveness principle: Identity, meaning, and the bounds of cultural relativity." *Personality and Social Psychology Review* 4:337-354.

Vinokur, Amiram, and Eugene Burnstein. 1978. "Depolarization of Attitudes in Groups." *Journal of Personality and Social Psychology* 36:872-885.

Wagner, Carl. 1982. "Allocation, Lehrer Models, and the Consensus of Probabilities." *Theory and Decision* 14:207-220.

Watts, D. J. 2003. *Six degrees: The Science of a connected age*. New York: Norton.

Watts, D. J., and P. S. Dodds. 2007. "Influentials, networks, and public opinion formation." *Journal of Consumer Research* 34:441-458.

Watts, D. J., and S. H. Strogatz. 1998. "Collective dynamics of 'small-world' networks." *Nature* 393:440-442.

Webber, Sheila Simsarian, and Lisa M. Donahue. 2001. "Impact of Highly and Less Job-related Diversity on Work Group Cohesion and Performance: A Meta-Analysis." *Journal of Management* 27:141-162.

Wegge, J., C. Roth, B. Neubach, K. H. Schmidt, and R. Kanfer. 2008. "Age and Gender Diversity as Determinants of Performance and Health in a Public Organization: The Role of Task Complexity and Group Size." *Journal of Applied Psychology* 93:1301-1313.

Weisbuch, Gèrard, Guillaume Deffuant, and Frédéric Amblard. 2005. "Persuasion Dynamics." *Physica A* 353:555-575.

Williams, Katherine Y., and Charles A. O'Reilly. 1998. "Demography and Diversity in Organizations. A Review of 40 Years of Research." *Research in Organizational Behavior* 20:77-140.

Wood, W. 2000. "Attitude change: Persuasion and social influence." *Annual Review of Psychology* 51:539-570.

Young, Peyton. 2001. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, N.J.: Princeton University Press.

ICS dissertation series

The ICS-series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen. Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS-series itself.

1) C. van Liere, (1990). *Lastige Leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen.* Amsterdam: Thesis Publishers.
2) Marco H.D. van Leeuwen, (1990). *Bijstand in Amsterdam, ca. 1800 - 1850. Armenzorg als beheersings- en overlevingsstrategie.* ICS dissertation, Utrecht.
3) I. Maas, (1990). *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?* Amsterdam: Thesis Publishers.
4) M.I. Broese van Groenou, (1991). *Gescheiden Netwerken. De relaties met vrienden en verwanten na echtscheiding.* Amsterdam: Thesis Publishers.
5) Jan M.M. van den Bos, (1991). *Dutch EC Policy Making. A Model-Guided Approach to Coordination and Negotiation.* Amsterdam: Thesis Publishers.
6) Karin Sanders, (1991). *Vrouwelijke Pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan.* Amsterdam: Thesis Publishers.
7) Sjerp de Vries, (1991). *Egoism, Altruism, and Social Justice. Theory and Experiments on Cooperation in Social Dilemmas.* Amsterdam: Thesis Publishers.
8) Ronald S. Batenburg, (1991). *Automatisering in bedrijf.* Amsterdam: Thesis Publishers.
9) Rudi Wielers, (1991). *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang.* Amsterdam: Thesis Publishers.
10) Gert P. Westert, (1991). *Verschillen in ziekenhuisgebruik.* ICS dissertation, Groningen.
11) Hanneke Hermsen, (1992). *Votes and Policy Preferences. Equilibria in Party Systems.* Amsterdam: Thesis Publishers.
12) Cora J.M. Maas, (1992). *Probleemleerlingen in het basisonderwijs.* Amsterdam: Thesis Publishers.
13) Ed A.W. Boxman, (1992). *Contacten en carrière. Een empirisch-theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities.* Amsterdam: Thesis Publishers.
14) Conny G.J. Taes, (1992). *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs.* Amsterdam: Thesis Publishers.
15) Peter van Roozendaal, (1992). *Cabinets in Multi-Party Democracies. The Effect of Dominant and Central Parties on Cabinet Composition and Durability.* Amsterdam: Thesis Publishers.
16) Marcel van Dam, (1992). *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid.* Amsterdam: Thesis Publishers.
17) Tanja van der Lippe, (1993). *Arbeidsverdeling tussen mannen en vrouwen.* Amsterdam: Thesis Publishers.
18) Marc A. Jacobs, (1993). *Software: Kopen of Kopiëren? Een sociaal-wetenschappelijk onderzoek onder PC-gebruikers.* Amsterdam: Thesis Publishers.
19) Peter van der Meer, (1993). *Verdringing op de Nederlandse arbeidsmarkt. Sector- en sekseverschillen.* Amsterdam: Thesis Publishers.
20) Gerbert Kraaykamp, (1993). *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag.* Amsterdam: Thesis Publishers.
21) Evelien Zeggelink, (1993). *Strangers into Friends. The Evolution of Friendship Networks Using an Individual Oriented Modeling Approach.* Amsterdam: Thesis Publishers.
22) Jaco Berveling, (1994). *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen.* Amsterdam: Thesis Publishers.

23) Wim Bernasco, (1994). *Coupled Careers. The Effects of Spouse's Resources on Success at Work.* Amsterdam: Thesis Publishers.
24) Liset van Dijk, (1994). *Choices in Child Care. The Distribution of Child Care Among Mothers, Fathers and Non-Parental Care Providers.* Amsterdam: Thesis Publishers.
25) Jos de Haan, (1994). *Research Groups in Dutch Sociology.* Amsterdam: Thesis Publishers.
26) K. Boahene, (1995). *Innovation Adoption as a Socio-Economic Process. The Case of the Ghanaian Cocoa Industry.* Amsterdam: Thesis Publishers.
27) Paul E.M. Ligthart, (1995). *Solidarity in Economic Transactions. An Experimental Study of Framing Effects in Bargaining and Contracting.* Amsterdam: Thesis Publishers.
28) Roger Th. A.J. Leenders, (1995). *Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure, and their Interdependence.* Amsterdam: Thesis Publishers.
29) Beate Völker, (1995). *Should Auld Acquaintance Be Forgot...? Institutions of Communism, the Transition to Capitalism and Personal Networks: the Case of East Germany.* Amsterdam: Thesis Publishers.
30) A. Cancrinus-Matthijsse, (1995). *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen.* Amsterdam: Thesis Publishers.
31) Nardi Steverink, (1996). *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen.* Amsterdam: Thesis Publishers.
32) Ellen Lindeman, (1996). *Participatie in vrijwilligerswerk.* Amsterdam: Thesis Publishers.
33) Chris Snijders, (1996). *Trust and Commitments.* Amsterdam: Thesis Publishers.
34) Koos Postma, (1996). *Changing Prejudice in Hungary. A Study on the Collapse of State Socialism and Its Impact on Prejudice Against Gypsies and Jews.* Amsterdam: Thesis Publishers.
35) Jooske T. van Busschbach, (1996). *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties.* Amsterdam: Thesis Publishers.
36) René Torenvlied, (1996). *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten.* Amsterdam: Thesis Publishers.
37) Andreas Flache, (1996). *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas.* Amsterdam: Thesis Publishers.
38) Kees van Veen, (1997). *Inside an Internal Labor Market: Formal Rules, Flexibility and Career Lines in a Dutch Manufacturing Company.* Amsterdam: Thesis Publishers.
39) Lucienne van Eijk, (1997). *Activity and Well-being in the Elderly.* Amsterdam: Thesis Publishers.
40) Róbert Gál, (1997). *Unreliability. Contract Discipline and Contract Governance under Economic Transition.* Amsterdam: Thesis Publishers.
41) Anne-Geerte van de Goor, (1997). *Effects of Regulation on Disability Duration.* ICS dissertation, Utrecht.
42) Boris Blumberg, (1997). *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus Empirisch-Theoretischer Perspektive.* ICS dissertation, Utrecht.
43) Marijke von Bergh, (1997). *Loopbanen van oudere werknemers.* Amsterdam: Thesis Publishers.
44) Anna Petra Nieboer, (1997). *Life-Events and Well-Being: A Prospective Study on Changes in Well-Being of Elderly People Due to a Serious Illness Event or Death of the Spouse.* Amsterdam: Thesis Publishers.
45) Jacques Niehof, (1997). *Resources and Social Reproduction: The Effects of Cultural and Material Resources on Educational and Occupational Careers in Industrial Nations at the End of the Twentieth Century.* ICS dissertation, Nijmegen.
46) Ariana Need, (1997). *The Kindred Vote. Individual and Family Effects of Social Class and Religion on Electoral Change in the Netherlands, 1956-1994.* ICS dissertation, Nijmegen.
47) Jim Allen, (1997). *Sector Composition and the Effect of Education on Wages: an International Comparison.* Amsterdam: Thesis Publishers.
48) Jack B.F. Hutten, (1998). *Workload and Provision of Care in General Practice. An Empirical Study of the Relation Between Workload of Dutch General Practitioners and the Content and Quality of their Care.* ICS dissertation, Utrecht.
49) Per B. Kropp, (1998). *Berufserfolg im Transformationsprozeß. Eine theoretisch-empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland.* ICS dissertation, Utrecht.
50) Maarten H.J. Wolbers, (1998). *Diploma-inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland.* ICS dissertation, Nijmegen.
51) Wilma Smeenk, (1998). *Opportunity and Marriage. The Impact of Individual Resources and Marriage Market Structure on First Marriage Timing and Partner Choice in the Netherlands.* ICS dissertation, Nijmegen.

52) Marinus Spreen, (1999). *Sampling Personal Network Structures: Statistical Inference in Ego-Graphs.* ICS dissertation, Groningen.
53) Vincent Buskens, (1999). *Social Networks and Trust.* ICS dissertation, Utrecht.
54) Susanne Rijken, (1999). *Educational Expansion and Status Attainment. A Cross-National and Over-Time Comparison.* ICS dissertation, Utrecht.
55) Mérove Gijsberts, (1999). *The Legitimation of Inequality in State-Socialist and Market Societies, 1987-1996.* ICS dissertation, Utrecht.
56) Gerhard G. Van de Bunt, (1999*). Friends by Choice. An Actor-Oriented Statistical Network Model for Friendship Networks Through Time.* ICS dissertation, Groningen.
57) Robert Thomson, (1999). *The Party Mandate: Election Pledges and Government Actions in the Netherlands, 1986-1998.* Amsterdam: Thela Thesis.
58) Corine Baarda, (1999). *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000.* ICS dissertation, Groningen.
59) Rafael Wittek, (1999). *Interdependence and Informal Control in Organizations.* ICS dissertation, Groningen.
60) Diane Payne, (1999*). Policy Making in the European Union: an Analysis of the Impact of the Reform of the Structural Funds in Ireland.* ICS dissertation, Groningen.
61)René Veenstra, (1999). *Leerlingen - Klassen - Scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs.* Amsterdam, Thela Thesis.
62) Marjolein Achterkamp, (1999). *Influence Strategies in Collective Decision Making. A Comparison of Two Models.* ICS dissertation, Groningen.
63) Peter Mühlau, (2000). *The Governance of the Employment Relation. A Relational Signaling Perspective.* ICS dissertation, Groningen.
64) Agnes Akkerman, (2000). *Verdeelde vakbeweging en stakingen. Concurrentie om leden.* ICS dissertation, Groningen.
65) Sandra van Thiel, (2000). *Quangocratization: Trends, Causes and Consequences.* ICS dissertation, Utrecht.
66) Rudi Turksema, (2000). *Supply of Day Care.* ICS dissertation, Utrecht.
67) Sylvia E. Korupp (2000*). Mothers and the Process of Social Stratification.* ICS dissertation, Utrecht.
68) Bernard A. Nijstad (2000). *How the Group Affects the Mind: Effects of Communication in Idea Generating Groups.* ICS dissertation, Utrecht.
69) Inge F. de Wolf (2000). *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers.* ICS dissertation, Utrecht.
70) Jan Kratzer (2001). *Communication and Performance: An Empirical Study in Innovation Teams.* ICS-dissertation, Groningen.
71) Madelon Kroneman (2001). *Healthcare Systems and Hospital Bed Use.* ICS/NIVEL-dissertation, Utrecht.
72) Herman van de Werfhorst (2001). *Field of Study and Social Inequality. Four Types of Educational Resources in the Process of Stratification in the Netherlands.* ICS-dissertation, Nijmegen.
73) Tamás Bartus (2001). *Social Capital and Earnings Inequalities. The Role of Informal Job Search in Hungary.* ICS-dissertation Groningen.
74) Hester Moerbeek (2001). *Friends and Foes in the Occupational Career. The Influence of Sweet and Sour Social Capital on the Labour Market.* ICS-dissertation, Nijmegen.
75) Marcel van Assen (2001). *Essays on Actor Perspectives in Exchange Networks and Social Dilemmas.* ICS-dissertation, Groningen.
76) Inge Sieben (2001). *Sibling Similarities and Social Stratification. The Impact of Family Background across Countries and Cohorts.* ICS-dissertation, Nijmegen.
77) Alinda van Bruggen (2001). *Individual Production of Social Well-Being. An Exploratory Study.* ICS-dissertation, Groningen.
78) Marcel Coenders (2001). *Nationalistic Attitudes and Ethnic Exclusionism in a Comparative Perspective: An Empirical Study of Attitudes Toward the Country and Ethnic Immigrants in 22 Countries.* ICS-dissertation, Nijmegen.
79) Marcel Lubbers (2001). *Exclusionistic Electorates. Extreme Right-Wing Voting in Western Europe.* ICS-dissertation, Nijmegen.
80) Uwe Matzat (2001). *Social Networks and Cooperation in Electronic Communities. A theoretical-empirical Analysis of Academic Communication and Internet Discussion Groups.* ICS-dissertation, Groningen.

81) Jacques P.G. Janssen (2002). *Do Opposites Attract Divorce? Dimensions of Mixed Marriage and the Risk of Divorce in the Netherlands,* ICS-dissertation, Nijmegen.
82) Miranda Jansen (2002). *Waardenoriëntaties en partnerrelaties. Een panelstudie naar wederzijdse invloeden,* ICS-dissertation, Utrecht.
83) Anne Rigt Poortman (2002). *Socioeconomic Causes and Consequences of Divorce.* ICS-dissertation, Utrecht.
84) Alexander Gattig (2002). *Intertemporal Decision Making*, ICS-dissertation, Groningen.
85) Gerrit Rooks (2002). *Contract en Conflict: Strategisch Management van Inkooptransacties,* ICS-dissertation, Utrecht.
86) Károly Takács (2002). *Social Networks and Intergroup Conflict.* ICS-dissertation, Groningen.
87) Thomas Gautschi (2002). *Trust and Exchange, Effects of Temporal Embeddedness and Network Embeddedness on Providing and Dividing a Surplus.* ICS-dissertation, Utrecht.
88) Hilde Bras (2002). *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850 – 1950.* Aksant Academic Publishers, Amsterdam.
89) Merijn Rengers (2002). *Economic Lives of Artists. Studies into Careers and the Labour Market in the Cultural Sector,* ICS-dissertation, Utrecht.
90) Annelies Kassenberg (2002). *Wat scholieren bindt. Sociale gemeenschap in scholen,* ICS-dissertation, Groningen
91) Marc Verboord (2003). *Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000.* ICS-dissertation, Utrecht.
92) Marcel van Egmond (2003). *Rain Falls on All of Us (but Some Manage to Get More Wet than Others): Political Context and Electoral Participation.* ICS-dissertation, Nijmegen.
93) Justine Horgan (2003). *High Performance Human Resource Management in Ireland and the Netherlands: Adoption and Effectiveness.* ICS-dissertation, Groningen.
94) Corine Hoeben (2003). *LETS' Be a Community. Community in Local Exchange Trading Systems.* ICS-dissertation, Groningen.
95) Christian Steglich (2003). *The Framing of Decision Situations. Automatic Goal Selection and Rational Goal Pursuit.* ICS-dissertation, Groningen.
96) Johan van Wilsem (2003). *Crime and Context. The Impact of Individual, Neighborhood, City and Country Characteristics on Victimization.* ICS-dissertation, Nijmegen.
97) Christiaan Monden (2003). *Education, Inequality and Health. The Impact of Partners and Life Course.* ICS-dissertation, Nijmegen.
98) Evelyn Hello (2003). *Educational Attainment and Ethnic Attitudes. How to Explain their Relationship.* ICS-dissertation, Nijmegen.
99) Marnix Croes en Peter Tammes (2004). *Gif laten wij niet voortbestaan. Een onderzoek naar de overlevingskansen van joden in de Nederlandse gemeenten, 1940-1945.* Aksant Academic Publishers, Amsterdam
100) Ineke Nagel (2004). *Cultuurdeelname in de levensloop.* ICS- dissertation, Utrecht.
101) Marieke van der Wal (2004). *Competencies to Participate in Life. Measurement and the Impact of School.* ICS-dissertation, Groningen.
102)Vivian Meertens (2004). *Depressive Symptoms in the General Population: a Multifactorial Social Approach.* ICS -dissertation, Nijmegen.
103) Hanneke Schuurmans (2004). *Promoting Well-Being in Frail Elderly People. Theory and Intervention.* ICS-dissertation, Groningen.
104) Javier Arregui (2004). *Negotiation in Legislative Decision-Making in the European Union.* ICS-dissertation, Groningen.
105) Tamar Fischer (2004). *Parental Divorce, Conflict and Resources. The Effects on Children's Behaviour Problems, Socioeconomic Attainment, and Transitions in the Demographic Career.* ICS-dissertation, Nijmegen.
106) René Bekkers (2004). *Giving and Volunteering in the Netherlands: Sociological and Psychological Perspectives.* ICS-dissertation, Utrecht.
107) Renée van der Hulst (2004). *Gender Differences in Workplace Authority: An Empirical Study on Social Networks.* ICS-dissertation, Groningen.
108) Rita Smaniotto (2004). *'You Scratch My Back and I Scratch Yours' Versus 'Love Thy Neighbour'. Two Proximate Mechanisms of Reciprocal Altruism.* ICS-dissertation, Groningen.
109) Maurice Gesthuizen (2004). *The Life-Course of the Low-Educated in the Netherlands: Social and Economic Risks.* ICS-dissertation, Nijmegen.

110) Carlijne Philips (2005). *Vakantiegemeenschappen. Kwalitatief en Kwantitatief Onderzoek naar Gelegenheid en Refreshergemeenschap tijdens de Vakantie.* ICS-dissertation, Groningen.
111) Esther de Ruijter (2005). *Household Outsourcing.* ICS-dissertation, Utrecht.
112) Frank van Tubergen (2005). *The Integration of Immigrants in Cross-National Perspective: Origin, Destination, and Community Effects.* ICS-dissertation, Utrecht.
113) Ferry Koster (2005). *For the Time Being. Accounting for Inconclusive Findings Concerning the Effects of Temporary Employment Relationships on Solidary Behavior of Employees.* ICS-dissertation, Groningen.
114) Carolien Klein Haarhuis (2005). *Promoting Anti-Corruption Reforms. Evaluating the Implementation of a World Bank Anti-Corruption Program in Seven African Countries (1999-2001).* ICS-dissertation, Utrecht.
115) Martin van der Gaag (2005). *Measurement of Individual Social Capital.* ICS-dissertation, Groningen.
116) Johan Hansen (2005). *Shaping Careers of Men and Women in Organizational Contexts.* ICS-dissertation, Utrecht.
117) Davide Barrera (2005). *Trust in Embedded Settings.* ICS-dissertation, Utrecht.
118) Mattijs Lambooij (2005). *Promoting Cooperation. Studies into the Effects of Long-Term and Short-Term Rewards on Cooperation of Employees.* ICS-dissertation, Utrecht.
119) Lotte Vermeij (2006). *What's Cooking? Cultural Boundaries among Dutch Teenagers of Different Ethnic Origins in the Context of School.* ICS-dissertation, Utrecht.
120) Mathilde Strating (2006). *Facing the Challenge of Rheumatoid Arthritis. A 13-year Prospective Study among Patients and Cross-Sectional Study among Their Partners.* ICS-dissertation, Groningen.
121) Jannes de Vries (2006). *Measurement Error in Family Background Variables: The Bias in the Intergenerational Transmission of Status, Cultural Consumption, Party Preference, and Religiosity.* ICS-dissertation, Nijmegen.
122) Stefan Thau (2006). *Workplace Deviance: Four Studies on Employee Motives and Self-Regulation.* ICS-dissertation, Groningen.
123) Mirjam Plantinga (2006). *Employee Motivation and Employee Performance in Child Care. The effects of the Introduction of Market Forces on Employees in the Dutch Child-Care Sector.* ICS-dissertation, Groningen.
124) Helga de Valk (2006). *Pathways into Adulthood. A Comparative Study on Family Life Transitions among Migrant and Dutch Youth.* ICS-dissertation, Utrecht.
125) Henrike Elzen (2006). *Self-Management for Chronically Ill Older People.* ICS-Dissertation, Groningen.
126) Ayse Güveli (2007). *New Social Classes within the Service Class in the Netherlands and Britain. Adjusting the EGP Class Schema for the Technocrats and the Social and Cultural Specialists.* ICS-dissertation, Nijmegen.
127) Willem-Jan Verhoeven (2007). *Income Attainment in Post-Communist Societies.* ICS-dissertation, Utrecht.
128) Marieke Voorpostel (2007). *Sibling support: The Exchange of Help among Brothers and Sisters in the Netherlands.* ICS-dissertation, Utrecht.
129) Jacob Dijkstra (2007). *The Effects of Externalities on Partner Choice and Payoffs in Exchange Networks.* ICS-dissertation, Groningen.
130) Patricia van Echtelt (2007). *Time-Greedy Employment Relationships: Four Studies on the Time Claims of Post-Fordist Work.* ICS-dissertation, Groningen.
131) Sonja Vogt (2007). *Heterogeneity in Social Dilemmas: The Case of Social Support.* ICS-dissertation, Utrecht.
132) Michael Schweinberger (2007). *Statistical Methods for Studying the Evolution of Networks and Behavior.* ICS-dissertation, Groningen.
133) István Back (2007). *Commitment and Evolution: Connecting Emotion and Reason in Long-term Relationships.* ICS-dissertation, Groningen.
134) Ruben van Gaalen (2007). *Solidarity and Ambivalence in Parent-Child Relationships.* ICS-dissertation, Utrecht.
135) Jan Reitsma (2007). *Religiosity and Solidarity – Dimensions and Relationships Disentangled and Tested.* ICS-dissertation, Nijmegen.
136) Jan Kornelis Dijkstra (2007). *Status and Affection among (Pre)Adolescents and Their Relation with Antisocial and Prosocial Behavior.* ICS-dissertation, Groningen.
137) Wouter van Gils (2007). *Full-time Working Couples in the Netherlands. Causes and Consequences.* ICS-dissertation, Nijmegen.
138) Djamila Schans (2007). *Ethnic Diversity in Intergenerational Solidarity.* ICS-dissertation, Utrecht.
139) Ruud van der Meulen (2007). *Brug over Woelig Water: Lidmaatschap van Sportverenigingen, Vriendschappen, Kennissenkringen en Veralgemeend Vertrouwen.* ICS-dissertation, Nijmegen.
140) Andrea Knecht (2008). *Friendship Selection and Friends' Influence. Dynamics of Networks and Actor Attributes in Early Adolescence.* ICS-dissertation, Utrecht.

141) Ingrid Doorten (2008). *The Division of Unpaid Work in the Household: A Stubborn Pattern?* ICS-dissertation, Utrecht.
142) Stijn Ruiter (2008). *Association in Context and Association as Context: Causes and Consequences of Voluntary Association Involvement.* ICS-dissertation, Nijmegen.
143) Janneke Joly (2008). *People on Our Minds: When Humanized Contexts Activate Social Norms.* ICS-dissertation, Groningen.
144) Margreet Frieling (2008). *'Joint production' als motor voor actief burgerschap in de buurt.* ICS-dissertion, Groningen.
145) Ellen Verbakel (2008*). The Partner as Resource or Restriction? Labour Market Careers of Husbands and Wives and the Consequences for Inequality Between Couples.* ICS-dissertation, Nijmegen.
146) Gijs van Houten (2008). *Beleidsuitvoering in gelaagde stelsels. De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg.* ICS-dissertation, Utrecht.
147) Eva Jaspers (2008*). Intolerance over Time. Macro and Micro Level Questions on Attitudes Towards Euthanasia, Homosexuality and Ethnic Minorities.* ICS-dissertation, Nijmegen.
148) Gijs Weijters (2008). *Youth delinquency in Dutch cities and schools: A multilevel approach.* ICS-dissertation, Nijmegen.
149) Jessica Pass (2009). *The Self in Social Rejection.* ICS-dissertation, Groningen.
150) Gerald Mollenhorst (2009). *Networks in Contexts. How Meeting Opportunities Affect Personal Relationships.* ICS-dissertation, Utrecht.
151) Tom van der Meer (2009). *States of freely associating citizens: comparative studies into the impact of state institutions on social, civic and political participation.* ICS-dissertation, Nijmegen.
152) Manuela Vieth (2009). *Commitments and Reciprocity in Trust Situations. Experimental Studies on Obligation, Indignation, and Self-Consistency.* ICS-dissertation, Utrecht.
153) Rense Corten (2009). *Co-evolution of social networks and behavior in social dilemmas: Theoretical and empirical perspectives.* ICS-dissertation, Utrecht.
154) Arieke J. Rijken (2009). *Happy families, high fertility? Childbearing choices in the context of family and partner relationships.* ICS-dissertation, Utrecht.
155) Jochem Tolsma (2009). *Ethnic Hostility among Ethnic Majority and Minority Groups in the Netherlands. An Investigation into the Impact of Social Mobility Experiences, the Local Living Environment and Educational Attainment on Ethnic Hostility.* ICS-dissertation, Nijmegen.
156) Freek Bucx (2009). *Linked Lives: Young Adults' Life Course and Relations With Parents.* ICS-dissertation, Utrecht.
157) Philip Wotschack (2009). *Household Governance and Time Allocation. Four studies on the combination of work and care.* ICS-dissertation, Groningen.
158) Nienke Moor (2009). *Explaining worldwide religious diversity. The relationship between subsistence technologies and ideas about the unknown in pre-industrial and (post-)industrial societies.* ICS-dissertation, Nijmegen.
159) Lieke ten Brummelhuis (2009). *Family matters at work. Depleting and enriching effects of employees' family lives on work outcomes.* ICS-dissertation, Utrecht.
160) Renske Keizer (2010). *Remaining childless. Causes and consequences from a life course perspective.* ICS-dissertation, Utrecht.
161) Miranda Sentse (2010). *Bridging Contexts: The interplay between family, child, and peers in explaining problem behavior in early adolescence.* ICS-dissertation, Groningen.
162) Nicole Tieben (2010). *Transitions, Tracks and Transformations. Social inequality in transitions into, through and out of secondary education in the Netherlands for cohorts born between 1914 and 1985.* ICS-dissertation, Nijmegen.
163) Birgit Pauksztat (2010). *Speaking up in organizations: Four studies on employee voice.* ICS-dissertation, Groningen.
164) Richard Zijdeman (2010). *Status attainment in the Netherlands, 1811-1941. Spatial and temporal variation before and during industrialization.* ICS-dissertation, Utrecht.
165) Rianne Kloosterman (2010). *Social background and children's educational careers. The primary and secondary effects of social background over transitions and over time in the Netherlands.* ICS-dissertation, Nijmegen.
166) Olav Aarts (2010). *Religious diversity and religious involvement. A study of religious markets in Western societies at the end of the twentieth century.* ICS-dissertation, Nijmegen.
167) Stephanie Wiesmann (2010). *24/7 Negotiation in couples transition to parenthood.* ICS-dissertation, Utrecht.

168) Borja Martinovic (2010). *Interethnic contacts: A dynamic analysis of interaction between immigrants and natives in Western countries.* ICS-dissertation, Utrecht.
169) Anne Roeters (2010). *Family life under pressure? Parents' paid work and the quantity and quality of parent-child and family time.* ICS-dissertation, Utrecht.
170) Jelle Sijtsema (2010). *Adolescent aggressive behavior: Status and stimulation goals in relation to the peer context.* ICS-dissertation, Groningen.
171) Michael Mäs (2010). *The Clustering Puzzle. Explaining clustering and polarization of opinions.* ICS-dissertation, Groningen.